Progress in Mathematics
319

Werner Ballmann
Christian Blohmann
Gerd Faltings
Peter Teichner
Don Zagier
Editors

Arbeitstagung Bonn 2013

In Memory of Friedrich Hirzebruch

Birkhäuser

Birkhäuser

Progress in Mathematics
Volume 319

More information about this series at http://www.springer.com/series/4848

Werner Ballmann • Christian Blohmann
Gerd Faltings • Peter Teichner • Don Zagier
Editors

Arbeitstagung Bonn 2013

In Memory of Friedrich Hirzebruch

Birkhäuser

Editors
Werner Ballmann
Max Planck Institute for Mathematics
Bonn, Germany

Christian Blohmann
Max Planck Institute for Mathematics
Bonn, Germany

Gerd Faltings
Max Planck Institute for Mathematics
Bonn, Germany

Peter Teichner
Max Planck Institute for Mathematics
Bonn, Germany

Don Zagier
Max Planck Institute for Mathematics
Bonn, Germany

ISSN 0743-1643 ISSN 2296-505X (electronic)
Progress in Mathematics
ISBN 978-3-319-43646-3 ISBN 978-3-319-43648-7 (eBook)
DOI 10.1007/978-3-319-43648-7

Library of Congress Control Number: 2016957869

Mathematics Subject Classification (2010): 00B20, 00B30

Printed on acid-free paper

This book is published under the trade name Birkhäuser, www.birkhauser-science.com
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The first Arbeitstagung was organized by Friedrich Hirzebruch in 1957. The six participants were Michael Atiyah, Hans Grauert, Alexander Grothendieck, Friedrich Hirzebruch, Nicolaas Kuiper, and Jacques Tits. Already at the first Arbeitstagung the principle was established which was to become its most distinctive feature: The program was not determined beforehand by the organizers, but during the meeting by all participants in an open discussion, albeit skillfully guided by Hirzebruch. The curiosity about who would be speaking and on what topic was part of the excitement of this conference series, which quickly grew in popularity, reaching well over 250 participants in later years. More importantly, the spontaneous proposal of talks made it possible to always cover the latest developments in mathematics. Many important results, such as the Atiyah-Singer index theorem, were first introduced to the larger mathematics community at the Arbeitstagung. Moreover, it meant that the conferences were not confined to a specific topic, representing over the years almost all fields of mathematics.

The Arbeitstagung 2013 was dedicated to the memory of Friedrich Hirzebruch, who passed away on May 27, 2012. This volume contains contributions from speakers and participants, covering a variety of topics from algebraic geometry, topology, differential geometry, operator theory, and representation theory. We hope that it still captures some of the unique character and spirit of Fritz Hirzebruch's original Arbeitstagung.

Bonn, Germany
June 2016

Werner Ballmann
Christian Blohmann
Gerd Faltings
Peter Teichner
Don Zagier

Contents

Contributors

Michael Atiyah School of Mathematics, JCMB, Edinburgh, UK

Anne-Marie Aubert Institut de Mathématiques de Jussieu – Paris Rive Gauche, U.M.R. 7586 du C.N.R.S., U.P.M.C., Paris, France

Werner Ballmann Max Planck Institute for Mathematics, Bonn, Germany

Christian Bär Institut für Mathematik, Universität Potsdam, Potsdam, Germany

Paul Baum Mathematics Department, Pennsylvania State University, University Park, PA, USA

Andrew Dancer Jesus College, Oxford, UK

Brent Doran Department of Mathematics, ETH Zürich, Zürich, Switzerland

Ben Elias Department of Mathematics, University of Oregon, Eugene, OR, USA

Joshua S. Friedman Department of Mathematics and Science, United States Merchant Marine Academy, Kings Point, NY, USA

Abhijit Gadde California Institute of Technology, Pasadena, CA, USA

Sergei Gukov California Institute of Technology, Pasadena, CA, USA

Nigel Hitchin Mathematical Institute, Radcliffe Observatory Quarter, Oxford, UK

Jay Jorgenson Department of Mathematics, The City College of New York, New York, NY, USA

Frances Kirwan Balliol College, Oxford, UK

Jürg Kramer Institut für Mathematik, Humboldt-Universität zu Berlin, Berlin, Germany

Laurentiu Maxim Department of Mathematics, University of Wisconsin-Madison, Madison, WI, USA

Dmitri I. Panyushev Institute for Information Transmission Problems of the R.A.S., Moscow, Russia

Independent University of Moscow, Moscow, Russia

Roger Plymen School of Mathematics, Southampton University, Southampton, UK

School of Mathematics, Manchester University, Manchester, UK

Pavel Putrov California Institute of Technology, Pasadena, CA, USA

Frank Quinn Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Simon Rose Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark

Morihiko Saito RIMS Kyoto University, Kyoto, Japan

Jörg Schürmann Mathematisches Institut, Universität Münster, Münster, Germany

Yan Soibelman Department of Mathematics, KSU, Manhattan, KS, USA

Maarten Solleveld Radboud Universiteit Nijmegen, Nijmegen, The Netherlands

Andrew Swann Department of Mathematics, Aarhus University, Aarhus, Denmark

CP^3-Origins, Centre of Excellence for Cosmology and Particle Physics Phenomenology, University of Southern Denmark, Odense, Denmark

Gerard van der Geer Korteweg-de Vries Instituut, Universiteit van Amsterdam, Amsterdam, The Netherlands

Marie-France Vignéras Institut de Mathematiques de Jussieu, Université de Paris 7, Paris, France

Geordie Williamson Max-Planck-Institut für Mathematik, Bonn, Germany

Noriko Yui Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada

The Hirzebruch Signature Theorem for Conical Metrics

Michael Atiyah

Dedicated to my friend Fritz Hirzebruch

1 Introduction

Exactly 60 years ago the young Fritz Hirzebruch came up with two spectacular theorems [H53, H54] which set the scene for the future development of algebraic geometry and topology. First there was his Signature theorem which gave an explicit topological formula for the Signature of the quadratic form on the middle cohomology of a compact oriented $4k$-dimensional manifold X

$$\mathrm{Sign}(X) = L_k(p_1, \ldots, p_k) \tag{1}$$

where p_i are the Pontryagin classes of X and L_k is an explicit polynomial with rational coefficients (see Sect. 2) of total weight k and so (by evaluation on the fundamental class of X) it gives a rational number (which by (1) is then an integer). The Signature of a non-degenerate real quadratic form is defined as $p - q$, where p is the number of $+$ signs and q the number of $-$ signs in a diagonalization.

This theorem, which rested on Thom's cobordism theory, was the jumping off point for the even more spectacular Hirzebruch–Riemann–Roch theorem (HRR). This gave an explicit cohomological formula for the Euler characteristic of the sheaf cohomology groups

$$\chi(X, W) = \sum_{q=0}^{n} (-1)^q \dim H^q(X, \mathscr{O}(W)) \tag{2}$$

This paper is based on a lecture given in Bonn in May 2013 at the Hirzebruch Memorial Conference but it incorporates significant improvements.

M. Atiyah (✉)
School of Mathematics, JCMB, The King's Buildings, Peter Guthrie Tait Road,
Edinburgh EH9 3FD
e-mail: M.Atiyah@ed.ac.uk

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_1

where X is a complex projective algebraic manifold of complex dimension n and W is an algebraic vector bundle on X. The formula is

$$\chi(X, W) = \{chW.T(X)\}[X] \tag{3}$$

where chW is the Chern character of W and T is the total Todd polynomial (see Sect. 2) in the Chern classes of X. Even the particular case where $W = 1$, the trivial line-bundle, was interesting since it identified two different definitions of the arithmetic genus. In this case (3) reduced to

$$\chi(X, 1) = T_n(X) \tag{4}$$

which is strikingly similar to (1), especially in view of the close formal connection between the L-polynomials and the Todd polynomials, which I will explain in Sect. 2.

There are situations where the presence of singularities means that formulae such as (1) or (4) do not hold and the difference between the two sets of equations called the "defect" of the singularity has been much studied by Hirzebruch and others. Of course, for this defect to be even defined, the two sides of the equation still have to make sense independently. The characteristic class side can be replaced by an integral expression, using the Chern–Weil representation for the Chern and Pontryagin classes, while the left side may still make sense in special situations.

There are three noteworthy cases when the defect can be well-defined:

1. X is a singular algebraic variety, so that sheaf cohomology and $\chi(X, W)$ are still defined.
2. X is a rational homology manifold (e.g. an orbifold) so that the Signature is still defined.
3. X is a manifold but one uses a Riemannian metric with singularities to compute the differential forms representing the characteristic classes.

Hirzebruch [H71] studied case 2 and showed an interesting connection with number theory via Dedekind sums. Hirzebruch also made a beautiful study of the cusps of Hilbert modular surfaces and computed the Signature defect (an example of case (1)) in terms of L-functions of real quadratic fields [H73]. He also conjectured a similar formula for all real number fields. This stimulated the work in [APS73] and led to a proof of the Hirzebruch conjecture in [ADS83]. The use of the symbol "L" in Hirzebruch Signature theorem, and in classical number theory, is fortuitous and almost prescient.

This paper can be viewed as falling under case (3). It concerns a Riemannian metric on X with conical singularities, of fixed angle, along a sub-manifold Y of co-dimension 2. As we will see in Sect. 3 the whole story can essentially be reduced to a local study of a cone in $\mathbb{R}^2$. It is elementary differential geometry/topology and involves no serious analysis (index theory).

The particular case of four-dimensional manifolds was studied in my joint paper [AL13] with Lebrun and the present paper was viewed as a natural generalization. In [AL13] several proofs were given and one version lent itself to generalization, but seemed to involve much more complicated formulae and so was postponed.

In fact, as a result of a key discussion with Don Zagier, the formulae turned not to be at all complicated. Further scrutiny led to new variants of the proof and the final version presented here is direct, elementary and transparent. I hope that Fritz would have liked this exposition.

Let me now outline the structure of the paper. Section 2 recalls the general notion of a genus attached to a power series

$$Q(x) = 1 + a_1 x + a_2 x^2 + \cdots \tag{5}$$

This was introduced by Hirzebruch and provided his key algebraic tool for studying multiplicative invariants such as the Todd genus and the Signature. I shall recall how this works for a fibration.

We will need to extend these standard results by allowing non-compact fibres and also conical singularities. The case we need will be for an $\mathbb{R}^2$-bundle over a compact manifold Y in which $\mathbb{R}^2$ has a special metric $\mathbb{R}^2(\beta)$ with a conical singularity of angle $2\pi\beta$ at the origin. The essential calculation is all for $\mathbb{R}^2(\beta)$, but exploiting its circular symmetry. We compute its first Chern form $c_1(\beta)^\sharp$ as an equivariant differential form (for the circle group).

This is then transferred to the normal bundle $Y(\beta)$ of Y in X with the metric $\mathbb{R}^2(\beta)$ in the fibre. In this way we get a formula (42) for the integral of the Q-class of $Y(\beta)$ along the fibres, for a particular connection.

In Sect. 4 we combine this with the tangential class coming from Y to compute the total integral $Q(Y(\beta))$, again for a particular connection, and we get a multiplicative formula as in Sect. 2. Finally, we argue that the same holds for the Levi-Civita connection.

We then extend this formula to the whole of X and derive our main result (Theorem 4.1) which computes the defect due to the conical metric for any Q-genus.

Applied with $Q = L$ we derive the explicit form of the signature defect (Theorem 4.2), which was our main aim.

2 Multiplicative Genera

I recall here the main ideas on multiplicative genera introduced by Hirzebruch. First consider a complex vector bundle W of rank k, over a space X whose total Chern class $c(W)$ is formally factorized as

$$c(W) = \Pi_{i=1}^{k}(1 + x_i) \tag{6}$$

so that the individual Chern classes $c_j(W)$ are the elementary symmetric functions of the x_i (which have formal dimension 2). If $Q(x)$ is a power series of the form (5) we then define

$$Q(W) = \Pi_{i=1}^{k} Q(x_i) \tag{7}$$

This is then a power series in the Chern classes of W

$$Q(W) = Q(c_1, c_2, \ldots) \tag{8}$$

which is easily seen to be independent of k (for large k) and terminates at the dimension of X. Moreover, from its definition, it is multiplicative in the sense that

$$Q(W_1 \oplus W_2) = Q(W_1)Q(W_2) \tag{9}$$

Two cases of special interest are

$$Q(x) = T(x) = \frac{x}{1 - e^{-x}} \tag{10}$$

$$Q(x) = L(x) = \frac{x}{\tanh x} \tag{11}$$

Note the elementary identity

$$T(2x) = L(x) + x \tag{12}$$

which expresses T as the sum of its even and odd parts.

The function (10) has the explicit expansion

$$T(x) = 1 + \frac{x}{2} + \sum_{m=1}^{\infty} \frac{(-1)^{m-1} B_m x^{2m}}{(2m)!} \tag{13}$$

which essentially defines the Bernoulli numbers B_m (in the even notation). They are positive rational numbers and the first few values are

$$B_1 = \frac{1}{6}, B_2 = \frac{1}{30}, \ldots\ldots\ldots, B_6 = \frac{691}{2730}, \ldots \tag{14}$$

showing large and intriguing denominators and numerators. Note that, because of (12), the infinite series in (13) with x replaced by $2x$ gives the expansion of $L(x)$.

It was Hirzebruch, through his work on the Todd genus and the Signature, who showed the arithmetical importance of the Bernoulli numbers in topology.

For an almost complex manifold we take W to be the tangent bundle of X, and get the total class $T(X)$ as a polynomial in the Chern classes $c_j(X)$:

$$T(X) = 1 + \frac{c_1}{2} + \frac{c_1^2 + c_2}{12} + \frac{c_1 c_2}{24} + \cdots \tag{15}$$

If $\dim X = 2n$, then (15) stops at $T_n(X)$. For compact X, HRR (as extended by the index theorem) asserts that $T_n(X)$, evaluated on the fundamental cycle is the arithmetic genus.

As noted, $L(x)$ in (10) is an even function of x and so we can define $L(X)$ for any real manifold as a polynomial in the Pontryagin classes $p_j(X)$, which are formally the elementary symmetric functions in the squares x_i^2

$$L(X) = \sum_{j=0}^{\infty} L_j(p_1, \ldots, p_j) \tag{16}$$

where $L_0 = 1, L_1 = \dfrac{p_1}{3}$ and L_j has dimension $4j$. If X is compact oriented and of dimension $4k$ the Hirzebruch Signature theorem asserts that

$$\mathrm{Sign}(X) = L_k(X) \tag{17}$$

where we evaluate on the fundamental cycle of X.

If X is a product $A \times Y$, with $\dim A = 4k$ $\dim Y = 4l$ the multiplicative property (9) shows that the Signature is multiplicative

$$L_{k+l}(A \times Y) = L_k(A) L_l(Y) \tag{18}$$

with a similar formula for the Todd genus.

A more interesting situation arises when X is not a product but a fibration $f : X \to Y$ with fibre A. It is now interesting to drop the restriction that both $\dim A$ and $\dim Y$ are divisible by 4. We only assume $\dim A$ and $\dim Y$ are even and that $\dim A + \dim Y$ is divisible by 4. For example, we can take $\dim A = \dim Y = 2$ and get interesting results as we shall see. For a fibration the tangent bundle $W(X)$ is then a sum

$$W(X) = W(f) \oplus f^* W(Y) \tag{19}$$

where $W(f)$ is the tangent bundle along the A-fibres. Then, for the L-genus, or more generally for any multiplicative genus Q, we have

$$Q(X) = Q(f).f^* Q(Y) \tag{20}$$

where $Q(f) = Q(W(f))$. As before $f^* Q(Y)$ is just a polynomial of degree l

$$f^* Q(Y) = \sum_{j=0}^{l} f^* Q_j(Y) \tag{21}$$

In the product (20) we therefore pick up, not only the obvious term as in (18), but also cross-terms

$$Q_{k+l-j}(f).f^{*}Q_j(Y) \tag{22}$$

To evaluate $Q(X)$ on the fundamental cycle of X we can first apply the push-forward $f_* : H^*(X) \to H^*(Y)$, (lowering dimensions by $\dim A$) and then evaluate on the fundamental cycle of Y

$$Q(X)[X] = \{f_*(Q(f))Q(Y)\}[Y] \tag{23}$$

The key term is thus seen to be

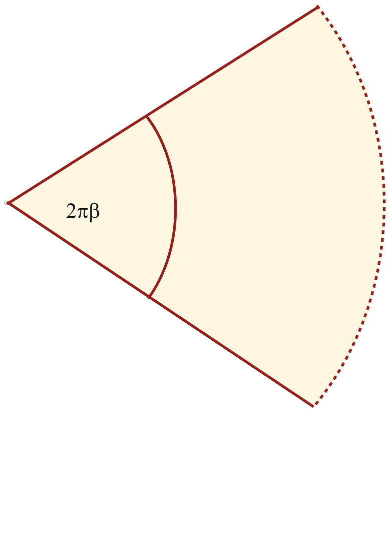
$$f_*Q(f) \in H^*(Y) \tag{24}$$

In the next section we shall need to apply formulae (23) and (24) to the situation where the fibre A is $\mathbb{R}^2$ with a special metric having a conical singularity, of angle $2\pi\beta$, at the origin and flat at infinity. There are therefore two new features, the non-compactness and the singular metric. We will show how to handle these in Sect. 4.

3 Cones

Our purpose in this section is to study two-dimensional cones, deriving explicit formulae which will be used to extend the multiplicative formulae (23) to allow conical singularities.

A cone in $\mathbb{R}^2$ of angle $2\pi\beta$, when slit open along a generator, gives a planar region as shown below



(25)

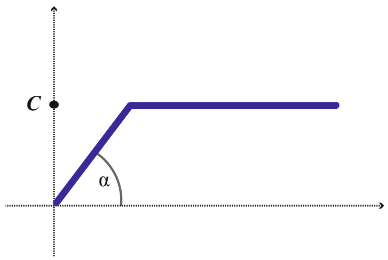
This shows that its natural metric is flat so that its tangent bundle (outside the vertex), with its metric connection, has zero-curvature but, near the vertex, there is non-trivial holonomy $2\pi(1-\beta)$ which can be interpreted as saying that the origin has a "distributional curvature" equal to $2\pi(1-\beta)$. Note that this is zero for $\beta = 1$, positive for $\beta < 1$ and negative for $\beta > 1$.

So far the picture would seem to imply that β should be restricted to the range $0 < \beta \leq 1$. However, in polar coordinates (r, θ), the metric on the cone takes the form

$$ds^2 = dr^2 + \beta^2 r^2 d\theta^2 \tag{26}$$

so that $\beta = 1$ is the standard metric in the plane. This formula for the metric shows that it makes sense for all positive β.

It will be convenient to modify the cone so that asymptotically it becomes a cylinder (of radius C). More formally we consider the graph



(27)

where $\sin\alpha = \beta$, smoothing it out at the break-point and then rotating it about at the x-axis. The resulting surface will be denoted by $\mathbb{R}^2(\beta)$. It depends on the actual smoothing chosen but the final formulae will depend only on β. Note that for $\beta = 1$ it is natural to take the smoothing to consist of a hemisphere with a cylinder attached, so that the (vertical) straight line segment in the graph is of zero length.

Topologically $\mathbb{R}^2(\beta)$ is independent of β and is just $\mathbb{R}^2$, but differentiably the vertex is (for $\beta \neq 1$) a singularity.

Consider now the tangent bundle of $\mathbb{R}^2(\beta)$ outside the vertex with its metric connection. At infinity and near the vertex it is flat (zero-curvature), the only curvature being in the annular region where the smoothing occurs. At infinity it is actually trivial but near the origin it has non-trivial holonomy $2\pi(1-\beta)$. Let c_1 be the first Chern form of this tangent bundle which is given by

$$c_1 = \frac{\kappa}{2\pi} w$$

where κ is the Gauss curvature and w the area 2-form of the metric. The large boundary circle (in the cylinder) is a geodesic while the total geodesic curvature round the small circle (near the vertex) is seen, from figure (25) to be $2\pi\beta$. So Gauss–Bonnet gives

$$\int_{\mathbb{R}^2(\beta)} c_1 = \beta \tag{28}$$

Note that the restriction $\beta < 1$ implied by diagram (25) is irrelevant for the argument leading to (28). The limit case $\beta = 1$ just follows from the formula for the area of a hemisphere.

We now want to exploit the rotational symmetry of the metric (26). We can now use G-equivariant differential forms on $\mathbb{R}^2$, leading to G-equivariant cohomology (where G is the circle group). A simple account of equivariant de Rham theory is given in [AB84, §4] and I shall use the results explained there, with minor changes of notation to fit in with this paper. Initially the metric plays no role. We let Ω^*_G be the G-invariant differential forms on $\mathbb{R}^2$ and we adjoin a degree 2 variable v to form the polynomial ring $\Omega^*_G[v]$. This has a natural differential operator $d^\sharp$ given by

$$d^\sharp a = da + i(\zeta)av$$

where $\zeta = \frac{\partial}{\partial\theta}$ is the vector field of the circle action and $i(\zeta)$ denotes the interior product. Then $(d^\sharp)^2 = 0$ and the resulting cohomology is

$$H^*_G(\mathbb{R}^2) \cong H^*_G(\text{point}) \cong \mathbb{R}[v] \tag{29}$$

By definition equivariant cohomology is the cohomology of the associated bundle over the classifying space

$$B_G = CP_\infty$$

which explains the origin of the polynomial ring $\mathbb{R}[v]$.

If $a \in \Omega^*_G$ is a G-invariant closed 2-form on $\mathbb{R}^2$, then we can extend it to an equivariantly closed form $a^\sharp$ by putting

$$a^\sharp = a - fv \tag{30}$$

when f is a function on $\mathbb{R}^2$ (unique up to a constant) with

$$df = i(\zeta)a \tag{31}$$

We want to apply this when

$$a = c_1(\beta) = \frac{\kappa\omega}{2\pi} \tag{32}$$

is the Chern form of the metric $\mathbb{R}^2(\beta)$. This is the place the metric enters.

Note that the area 2-form ω defines a symplectic structure preserved by the circle action. Let u be the Hamiltonian of the action, so that

$$\omega = du_\wedge d\theta$$

Near the origin, where the metric is Euclidean, $u = \beta\frac{r^2}{2}$, while at infinity $u \sim Cdr_\wedge d\theta$. Hence

$$i(\zeta)c_1(\beta) = -\frac{\kappa}{2\pi} \tag{33}$$

The function f of (31), with the choice (32) of a and the resulting formula (33), is then defined by the indefinite integral

$$f(r) = \int_{\frac{r^2}{2}}^{\infty} \kappa du \tag{34}$$

The choice of ∞ as the upper limit normalizes the arbitrary constant by making $f(\infty) = 0$. Note that the metric $\mathbb{R}^2(\beta)$ is flat for all large r, so $f(\infty)$ is just the value of $f(r)$ for large r.

To sum up we have extended the Chern form $c_1(\beta)$ on $\mathbb{R}^2(\beta)$ to an equivariantly closed form $c_1(\beta)^\sharp$ defined by

$$c_1(\beta)^\sharp = c_1(\beta) - f(r)v \tag{35}$$

Since $f(r)$=0 for all large r and since $c_1(\beta)$ vanishes near the vertex we can view $c_1(\beta)^\sharp$ as an equivariant form on the whole of $\mathbb{R}^2$ with **compact support**. In particular we can restrict it to the origin.

Note that the value $f(0)$ is given, using (28), by the definite integral

$$f(0) = \int_0^\infty \kappa dt = \int_{\mathbb{R}^2(\beta)} \frac{\kappa}{2\pi}\omega = \beta \tag{36}$$

where the passage from the double integral to the single integral can be carried out on any small circle $r = \varepsilon$ (since the metric is flat near the origin).

Thus our equivariant Chern form $c_1(\beta)^\sharp$, given by (35), restricts to the origin to give the class βv

$$c_1(\beta)^\sharp|_0 = \beta v \tag{37}$$

Now let us return to the equivariant cohomology. Besides $H_G^*(\mathbb{R}^2)$ we can consider the equivariant cohomology with compact support $H_{G.c}^*(\mathbb{R}^2)$. This is given by the same complex $(\Omega_G^*[v], d^\sharp)$ but now requiring that we restrict to forms in Ω_G^* with compact support. This gives the equivariant cohomology of the "Thom space" of $\mathbb{R}^2(\beta)$ which is a free module over the equivariant cohomology of $\mathbb{R}^2$ with one generator λ which is characterized by its integral over $\mathbb{R}^2$ being equal to 1. The restriction of λ to the origin is the class v.

It follows from (37) that

$$[c_1(\beta)^\sharp] = \beta\lambda \tag{38}$$

in the compactly supported equivariant cohomology. With this crucial formula we now consider the general operation of integrating over $\mathbb{R}^2(\beta)$: the push-forward

$$\int : H^*_{G,c}(\mathbb{R}^2) \to H^*_G(\text{point}) = \mathbb{R}[v]$$

For any class a with compact support we have, for $n \geq 1$,

$$[a^n] = [a].[a^{n-1}] = [a].[a(0)^{n-1}]$$

where $a(0) \in \mathbb{R}[v]$ is the value at the origin. Hence integration over $\mathbb{R}^2$ gives

$$\int a^n = a(0)^{n-1} \int a$$

Applied to $a = c_1(\beta)^\sharp$ and using (37) and (38) we get

$$\int (c_1(\beta)^\sharp)^n = \beta^{n-1} v^{n-1} \beta = \beta^n v^{n-1} \tag{39}$$

Now let $Q(x) = 1 + a_1 x + a_2 x^2 + \cdots$ be any power series, then (39) leads to

$$\int Q(c_1(\beta)^\sharp) = \frac{Q(\beta v) - 1}{v} \tag{40}$$

This is our key formula, expressed in terms of equivariant cohomology. By the standard Chern–Weil process it can now be transported immediately into a formula computing the expression (24) for the case when $A = \mathbb{R}^2(\beta)$ and $f : Y(\beta) \to Y$ is the bundle projection associated with a principal circle bundle over Y. The class v gets converted into the first Chern class y of P. In the notation of Sect. 2, we have

$$f_* Q(f) = \frac{Q(\beta y) - 1}{y} \tag{41}$$

Note. In this calculation we have used the Chern form defined by a particular (direct sum) connection. In Sect. 4 we will see that the same results hold for the Levi-Civita connection.

4 Globalizing the Argument

In the previous section we gave a detailed analysis of the tangent bundle along the fibres of the map

$$f : Y(\beta) \to Y$$

where the fibres are the modified planes $\mathbb{R}^2(\beta)$ with a conical singularity of angle $2\pi\beta$ at the origin. For any multiplicative genus Q, given by a power series (5), we computed

$$f_* Q(f) \in H^*(Y)$$

deriving the formula (41).

We now need to bring in the factor $Q(Y)$ coming from the base and then show how to derive the Q-defect due to Y. There are three technical problems to deal with

A. the conical singularity
B. the non-compactness of $Y(\beta)$ at infinity (far from the singularity)
C. the fact that the Levi-Civita connection of $Y(\beta)$ differs from the connection of the direct sum (see [O66]).

Consider problem (A). Since this is a local question we may replace Y by a small ball and also restrict to the purely conical region.[1] For simplicity we still use the same notation $Y(\beta)$.

Topologically all these $Y(\beta)$, being homeomorphic to $Y(1)$, are just products $Y \times D$ where D is the punctured 2-disc. The universal cover $\hat{Y}(1)$ is thus the product $Y \times H$ where H is the open half-plane (r, θ) with $r > 0$. The action of the circle $\mathbb{G}$ on $Y(1)$ lifts to θ-translation on H. Dividing by the subgroup of $\mathbb{G}$ generated by $\theta = 2\pi\beta$ we recover the cone $\mathbb{R}(\beta)$.

Now consider the Levi-Civita connection Δ on $Y(1)$ using the metric on Y and the connection on its normal bundle in X. This is smooth along the zero-section so its lift to $\hat{Y}(1)$ extends to the closed half-plane bundle $Y \times \bar{H}$ given by taking $r \geq 0$. The θ-translation on $Y \times \bar{H}$ is an isometry and so preserves the Levi-Civita connection.

When we divide by 2π the radial limits along the fibres lead to smoothness at the origin.

If we divide by $2\pi\beta$, for β not an integer, these radial limits will depend on the radius only by an orthogonal transformation. Thus the Riemannian curvature and the Pontryagin forms will remain bounded and all integrals will converge.

The same remarks apply to all the connections ∇_t of the 1-parameter family

$$\nabla_t = t\nabla + (1-t)\nabla'$$

where ∇' is the connection on $Y(1)$ given by the direct sum of the base and fibre connections (the one we used in Sect. 3).

[1] For $\beta = 1$ there is no purely conical region if we take the hemispherical smoothing, but in this case there is no singularity.

Thus, as far as the behaviour near the conical singularity goes, things behave as in the compact case and the integrals of Pontryagin forms are independent of the connection (in our family ∇_t). This deals with both problem (A) and problem (C).

Problem (B) is disposed of for our purposes, by integrating, not the Pontryagin forms of $Y(\beta)$, but the difference between these forms and those of $Y(1)$. Since $Y(\beta)$ and $Y(1)$ are isometric at infinity the difference has compact support and so presents no problems, whichever connection we use.

To sum up we have shown that the multiplicative formula (20) holds for the difference $Q(Y(1)) - Q(Y(\beta))$, and hence for the Q-defect

$$\delta Q(\beta) = \int_X Q(X) - \int_{X(\beta)} Q(X(\beta)) \tag{42}$$

We now use formula (41), both for $Y(\beta)$ and for $Y(1)$. Taking the difference as in (42) we end up with our main result:

Theorem 4.1. *The defect $\delta(Q(\beta))$ for the Q-genus of a manifold X with conical metric $\rho(\beta)$, of angle $2\pi(\beta)$, along the co-dimension two sub-manifold Y is given by*

$$\delta Q((\beta)) = \left\{ \frac{Q(y) - Q(\beta y)}{y} Q(Y) \right\} [Y]$$

where y is the class dual to Y in X.

Note. We gave the proof of this for orientable Y, but as we saw the essential calculation is local near Y. Hence if Y is non-orientable we can lift the integrals to the double covering. This means that the theorem continues to hold but y is now interpreted as a cohomology class (in the neighbourhood of Y) with twisted coefficients.

Our original integral was for $Q = L$ giving the Hirzebruch formula for the signature. Using the power series expansion (derived from (12) and (13)).

$$L(x) = 1 + \sum_{m=1}^{\infty} \frac{(-1)^{m-1} B_m 2^{2m} x^{2m}}{(2m)!} \tag{43}$$

we obtain

Theorem 4.2. *Let $X(\beta)$ be a compact oriented $4k$-dimensional Riemannian manifold with a conical singularity of angle $2\pi\beta$ along a co-dimension two sub-manifold Y, then the Signature defect $\delta(\beta)$ is given by*

$$\delta(\beta) = \left\{ \sum_{m=1}^{k} \frac{(-1)^{m-1} B_m (1 - \beta^{2m}) 2^{2m} y^{2m-1} L_{k-m}}{(2m)!} (Y) \right\} [Y]$$

where y is the first Chern class of the normal bundle of Y in X.

Notes.

1. As noted before, it Y is non-orientable, y has to be interpreted as a "twisted" class, with local coefficients.
2. The surprising thing about the formula of Theorem 4.2 is not the apparent complication with many terms but the simplicity of its dependence on the parameter β.
3. The case $\beta = 1/q$, where q is an integer, arises naturally from the situation of a q-fold branched cover $X' \to X$, branched along Y. A metric on X' invariant under the cyclic group action induces a metric on X with conical singularities along Y of angle $\dfrac{2\pi}{q}$. This is then a special case of the Signature defect studied by Hirzebruch, except that our definition of the defect arising from an integral over X, differs from that of Hirzebruch by a factor q.
4. The strict definition of conical adopted here can clearly be relaxed.

Let us consider the first two cases of Theorem 4.2, where $k = 1$ or 2.

For $k = 1, \dim X = 4$ and Theorem 4.2 reduces to the simple expression

$$\delta(\beta) = \frac{(1-\beta^2)}{3} Y^2 \tag{44}$$

where Y^2 is the self-intersection number of Y. This is the formula given in [AL13].

For $k = 2, \dim X = 8$, and the series in Theorem 4.2 now has two terms, so that

$$\delta(\beta) = \left\{ \frac{(1-\beta^2) y p_1}{9} - \frac{(1-\beta^4) y^3}{45} \right\} [Y] \tag{45}$$

where p_1 is the first Pontryagin class of Y.

As an example and a simple check, take $X = P_4(C), Y = Q_3$ a quadric threefold, $\beta = \frac{1}{2}$, so that the branched double covering $X' \to X$ has $X' = Q_4$. The total Chern class of Q_3 is

$$c(Q_3) = (1+x)^5 (1+2x)^{-1} = 1 + 3x + 4x^2 + \cdots \tag{46}$$

so that $p_1 = c_1^2 - 2c_2 = x^2$. Since $y = 2x$, formula (45) gives

$$\begin{aligned} \delta\left(\frac{1}{2}\right) &= \frac{3.2x.x^2}{4.9} - \frac{15.8x^3}{16.45} \\ &= \frac{1}{6}(x^3 - x^3) = 0 \end{aligned}$$

On the other hand, for a q-fold branched covering $X' \to X$ the defect is

$$\delta = \mathrm{Sign} X - \frac{1}{q} \mathrm{Sign} X' \tag{47}$$

In our case $q = 2, \ \ \mathrm{Sign} Q_4 = 2, \ \ \mathrm{Sign} P_4 = 1$, so the calculation checks.

In fact, following the original method of Hirzebruch based on Cauchy residue calculations, it is amusing to verify Theorem 4.2 for all the double coverings

$$Q_{2n} \to P_{2n} \qquad Y = Q_{2n-1}$$

Note first that, generalizing (46), we have

$$L(Q_{2n-1}) = L(1+x)^{n+1}L(1+2x)^{-1} = \left(\frac{x}{\tanh x}\right)^{2n+1}\frac{\tanh 2x}{2x}$$

Theorem 4.2, in the form of Theorem 4.1 for $Q = L$, but without expanding the power series, gives

$$\begin{aligned}\delta &= \left\{\frac{L(2x)-L(x)}{2x}L(Q_{2n-1})\right\}[Q_{2n-1}]\\ &= \text{coefficient of} \quad x^{2n-1} \quad \text{in}\\ &\quad \frac{1}{2x}\left\{\left(\frac{x}{\tanh x}\right)^{2n+1} - \frac{x^{2n+1}}{2(\tanh x)^{2n+2}}\tanh 2x\right\}\\ &= \text{Res}\frac{1}{2}\left\{\left(\frac{1}{\tanh x}\right)^{2n+1} - \frac{\tanh 2x}{2(\tanh x)^{2n+2}}\right\}dx\end{aligned}$$

where the residue is taken at $x = 0$. Now substitute $t = \tanh x, dx = dt/1-t^2$, so that $\tanh 2x = \dfrac{2t}{1+t^2}$ and

$$\begin{aligned}\delta &= \text{Res}\frac{1}{2}\left(\frac{1}{t^{2n+1}} - \frac{1}{t^{2n+2}}\cdot\frac{t}{1+t^2}\right)\frac{dt}{1-t^2}\\ &= \text{coefficient of} \quad t^{2n} \quad \text{in}\\ &\quad \left\{(1-t^2)^{-1} - (1-t^4)^{-1}\right\}\\ &= \begin{cases} 0 & \text{if } n \text{ is even}\\ 1 & \text{if } n \text{ is odd}\end{cases}\end{aligned}$$

Using (47) this checks with the well-known fact that

$$\text{Sign}Q_{2n} = \begin{cases} 2 & \text{if } n \text{ is even}\\ 0 & \text{if } n \text{ is odd}\end{cases}$$

Cases when $\beta = \dfrac{1}{q}$ with q integral are at least locally given by branched covers and the G-Signature theorem [AS68] gives the explicit form of the Lefschetz number

$L(g)$ (for $g \neq 1$ in the cyclic group of order q). Summing over all $g \neq 1$ then gives the defect and this was the starting point for the proof in [AL13], but which proceeded from inverse integers $\frac{1}{q}$, to rational $\frac{p}{q}$ and then by continuity to all real β. It also used more general machinery (index theory, cobordism) as well as complicated trigonometric sums. All this has been avoided in our present version.

Our method also shows that there is nothing special about the L-genus, since Theorem 4.1 works for any genus Q.

For example, we could take $Q = \hat{A}$, which corresponds to the index of the Dirac operator and obtain the defect formula for $\hat{A}$, which is very close to the Todd genus.

Finally I should comment on the limiting case $\beta = 0$. If we consider $\beta \to 0$ a cone with fixed base will tend to a cylinder. Thus $\beta = 0$ corresponds to non-compact manifolds having a complete cylindrical type metric based on Y. For an appropriate class of complete metrics we would then expect the formula for the Signature in Theorem 4.2 to hold with $\beta = 0$. In [AL13] interesting examples of this type occurred and were in fact the primary source of the problem.

In addition to Don Zagier I acknowledge with thanks the advice and comments of Nigel Hitchin, Claude Lebrun and Jean-Michel Bismut.

References

[AB84] M.F. Atiyah, R. Bott, The moment map and equivariant cohomology. Topology **23**(1), 1–28 (1984)

[AL13] M.F. Atiyah, C. Lebrun, Curvature, cones and characteristic numbers. Math. Proc. Camb. Phil. Soc. **155**, 13–37 (2013)

[AS68] M.F. Atiyah, I.M. Singer, The index of elliptic operators III. Ann. Math. **87**, 546–604 (1968)

[ADS83] M.F. Atiyah, H. Donnelly, I.M. Singer, Eta invariants, signature defects of cusps and values of L-functions. Ann. Math. **118**, 131–77 (1983)

[APS73] M.F. Atiyah, V.K. Patodi, I.M. Singer, Spectral asymmetry and Riemannian geometry. Bull. Lond. Math. Soc. **5**, 229–234 (1973)

[H53] F. Hirzebruch, On Steenrod's reduced powers, the index of inertia, and the Todd genus. Proc. Nat. Acad. Sci. USA **39**, 951–956 (1953)

[H54] F. Hirzebruch, Arithmetic genera and the theorem of Riemann-Roch for algebraic varieties. Proc. Nat. Acad. Sci. USA **40**, 110–114 (1954)

[H71] F. Hirzebruch, The signature theorem, reminiscences and recreations, in prospects in mathematics. Ann. Math. Study **70**, 3–31 (1971)

[H73] F. Hirzebruch, Hilbert modular surfaces. Enseign. Math. **19**, 183–281 (1973)

[O66] B. O'Neill, The fundamental equations of a submersion. Mich. Math. J. **13**, 459–469 (1966)

Depth and the Local Langlands Correspondence

Anne-Marie Aubert, Paul Baum, Roger Plymen, and Maarten Solleveld

Abstract Let G be an inner form of a general linear group over a non-archimedean local field. We prove that the local Langlands correspondence for G preserves depths. We also show that the local Langlands correspondence for inner forms of special linear groups preserves the depths of essentially tame Langlands parameters.

Keywords Representation theory • p-adic group • Local Langlands program • Division algebra

2010 *Mathematics Subject Classification.* 20G25, 22E50.

1 Introduction

Let F be a non-archimedean local field and let G be a connected reductive group over F. Let $\Phi(G)$ denote the collection of equivalence classes of Langlands parameters for G, and $\mathrm{Irr}(G)$ the set of (isomorphism classes of) irreducible smooth

A.-M. Aubert
Institut de Mathématiques de Jussieu – Paris Rive Gauche, U.M.R. 7586 du C.N.R.S., U.P.M.C., 4 place Jussieu, 75005 Paris, France
e-mail: anne-marie.aubert@imj-prg.fr

P. Baum (✉)
Mathematics Department, Pennsylvania State University, University Park, PA 16802, USA
e-mail: baum@math.psu.edu

R. Plymen
School of Mathematics, Southampton University, Southampton SO17 1BJ, UK

School of Mathematics, Manchester University, Manchester M13 9PL, UK
e-mail: r.j.plymen@soton.ac.uk; plymen@manchester.ac.uk

M. Solleveld
Radboud Universiteit Nijmegen, Heyendaalseweg 135, 6525AJ Nijmegen, The Netherlands
e-mail: m.solleveld@science.ru.nl

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_2

G-representations. A central role in the representation theory of such groups is played by the local Langlands correspondence (LLC). It is supposed to be a map

$$\mathrm{Irr}(G) \to \Phi(G)$$

that enjoys several naturality properties [Bor, Vog]. The LLC should preserve interesting arithmetic information, like local L-functions and ϵ-factors. A lesser-known invariant that makes sense on both sides of the LLC is *depth*.

The depth of a Langlands parameter ϕ is easy to define. For $r \in \mathbb{R}_{\geq 0}$ let $\mathrm{Gal}(F_s/F)^r$ be the r-th ramification subgroup of the absolute Galois group of F. Then the depth of ϕ is the smallest number $d(\phi) \geq 0$ such that ϕ is trivial on $\mathrm{Gal}(F_s/F)^r$ for all $r > d(\phi)$.

The depth $d(\pi)$ of an irreducible G-representation π was defined by Moy and Prasad [MoPr1, MoPr2], in terms of filtrations $P_{x,r}(r \in \mathbb{R}_{\geq 0})$ of the parahoric subgroups $P_x \subset G$. On the basis of several examples (see below) it is reasonable to expect that for many Langlands parameters $\phi \in \Phi(G)$ with L-packet $\Pi_\phi(G) \subset \mathrm{Irr}(G)$ one has

$$d(\phi) = d(\pi) \quad \text{for all } \pi \in \Pi_\phi(G). \tag{1}$$

This relation would be useful for several reasons. Firstly, it allows one to employ some counting arguments in the local Langlands correspondence, because (up to unramified twists) there are only finitely many irreducible representations and Langlands parameters whose depth is at most a specified upper bound.

Secondly, it would be a step towards a more explicit LLC. One can try to determine the groups $P_{x,r}/P_{x,r+\epsilon}$ ($\epsilon > 0$ small) and their representations explicitly, and to match them with representations of $\mathrm{Gal}(F_s/F)/\mathrm{Gal}(F_s/F)^{r+\epsilon}$.

Thirdly, one can use (1) as a working hypothesis when trying to establish a local Langlands correspondence, to determine whether or not two irreducible representations stand a chance of belonging to the same L-packet.

The most basic case of depth preservation concerns Langlands parameters $\phi \in \Phi(G)$ that are trivial both on the inertia group $\mathbf{I}_F$ and on $\mathrm{SL}_2(\mathbb{C})$. These can be regarded as Langlands parameters of negative depth. Such a ϕ is only relevant for G if G is quasi-split and splits over an unramified extension of F. In that case one can say that an irreducible G-representation has negative depth if it possesses a nonzero vector fixed by a hyperspecial maximal compact subgroup. The Satake isomorphism shows how to each such representation one can associate (in a natural way) a Langlands parameter of the above kind.

The G-representations of depth zero have been subjected to ample study, see, for example, [GSZ, Mor, DBRe, Mœ]. According to Moy–Prasad, an irreducible representation has depth zero if and only if it has nonzero vectors fixed by the pro-unipotent radical of some parahoric subgroup of G. This includes Iwahori-spherical representations and Lusztig's unipotent representations [Lus1, Lus2]. A Langlands parameter has depth zero if and only if it is trivial on the wild inertia subgroup of the absolute Galois group of F. For depth zero the equality (1) is conjectured, and proven in certain cases, in [DBRe]. It fits very well with the aforementioned work of Lusztig.

In positive depth there is the result of Yu [Yu2, §7.10], who proved (1) for unramified tori. For $\mathrm{GL}_n(F)$, (1) was claimed in [Yu1, §2.3.6] and proved in [ABPS2, Proposition 4.5]. For $\mathrm{GSp}_4(F)$, (1) is proved in [Gan, §10]. We refer to [GrRe, Ree, ReYu] for some interesting examples of positive depth Langlands parameters and supercuspidal representations. Most of these examples satisfy (1), but in [ReYu, §7.4–7.5] some particular cases are mentioned in which (1) does not hold. All these counterexamples appear in small residual characteristics. So it remains to be seen in which generality the local Langlands correspondence will preserve depths.

In this paper we will prove that the local Langlands correspondence preserves depth for the inner forms of $\mathrm{GL}_n(F)$. In a few non-split cases, this was done before in [LaRa]. For inner forms of $\mathrm{SL}_n(F)$, we will prove an inequality between depths, which becomes an equality if the Langlands parameter is essentially tame in the sense that it maps the wild inertia group to a maximal torus of $\mathrm{PGL}_n(\mathbb{C})$. Every Langlands parameter for an inner form of $\mathrm{SL}_n(F)$ is essentially tame if the residual characteristic of F does not divide n.

Let D be a division algebra with centre F, of dimension d^2 over F. Then $G = \mathrm{GL}_m(D)$ is an inner form of $\mathrm{GL}_n(F)$ with $n = dm$. There is a reduced norm map $\mathrm{Nrd}\colon \mathrm{GL}_m(D) \to F^\times$ and the derived group $G_{\mathrm{der}} := \ker(\mathrm{Nrd} : G \to F^\times)$ is an inner form of $\mathrm{SL}_n(F)$. Every inner form of $\mathrm{GL}_n(F)$ or $\mathrm{SL}_n(F)$ is isomorphic to one of this kind.

The main steps in the proof of our depth-preservation theorem are

- With the Langlands classification one reduces the problem to essentially square-integrable representations and elliptic Langlands parameters.
- Express the depth in terms of ϵ-factors and conductors. This is a technical step which involves detailed knowledge of the representation theory of G. Here it is convenient to use an alternative but equivalent version of depth, the normalized level of an irreducible G-representation.
- Show that the Jacquet–Langlands correspondence for $G = \mathrm{GL}_m(D)$ preserves ϵ-factors. Since the LLC for $\mathrm{GL}_m(D)$ is defined as a composition of the Jacquet–Langlands correspondence with the LLC for $\mathrm{GL}_n(F)$ and the latter is known to preserve ϵ-factors, this proves depth preservation for inner forms of $\mathrm{GL}_n(F)$.
- Relate the depth for G_{der} to depth for G. For irreducible representations nothing changes, but for Langlands parameters the depth can decrease if one replaces the dual group $\mathrm{GL}_n(\mathbb{C})$ by $\mathrm{PGL}_n(\mathbb{C})$. Using several properties of the Artin reciprocity map, we show that such a decrease in depth cannot occur if the Langlands parameter is essentially tame.

This paper develops results presented by the second author in a lecture at the 2013 Arbeitstagung.

Acknowledgements The authors wish to thank Vincent Sécherre for some helpful explanations on the construction of simple types for $\mathrm{GL}_m(D)$, Mark Reeder for pointing out some examples where the depth is not preserved, Wilhelm Zink for explaining properties of the Artin reciprocity map and Paul Broussous for providing the reference [BaBr], which has allowed a substantial simplification of the proof of Proposition 2.6 from a previous version. The second author "Paul Baum" was partially supported by NSF grant DMS-1200475.

2 The Local Langlands Correspondence for Inner Forms of $\mathrm{GL}_n(F)$

2.1 The Statement of the Correspondence

The local Langlands correspondence for supercuspidal representations of $\mathrm{GL}_n(F)$ was established in the important papers [LRS, HaTa, Hen2]. Together with the Jacquet–Langlands correspondence this provides the LLC for essentially square-integrable representations of inner forms $G = \mathrm{GL}_m(D)$ of $\mathrm{GL}_n(F)$. It is extended to all irreducible G-representations via the Zelevinsky classification [Zel, DKV], see [HiSa, ABPS1]. For these groups every L-packet is a singleton and the LLC is a canonical bijective map

$$\mathrm{rec}_{D,m} : \mathrm{Irr}(\mathrm{GL}_m(D)) \to \Phi(\mathrm{GL}_m(D)). \tag{2}$$

A remarkable aspect of Langlands' conjectures [Vog] is that it is better to consider not just one reductive group at a time, but all inner forms of a given group simultaneously. Inner forms share the same Langlands dual group, so in (2) the right-hand side is the same for all inner forms G of the given group. Then one can turn (2) into a bijection by defining a suitable equivalence relation on the set of inner forms and taking the corresponding union of the sets $\mathrm{Irr}(G)$ on the left-hand side (see Theorem 2.1 below).

We define the equivalence classes of such inner forms to be in bijection with the isomorphism classes of central simple F-algebras of dimension n^2 via $\mathrm{M}_m(D) \mapsto \mathrm{GL}_m(D)$, respectively, $\mathrm{M}_m(D) \mapsto \mathrm{GL}_m(D)_{\mathrm{der}}$.

As Langlands dual group of G we take $\mathrm{GL}_n(\mathbb{C})$. To deal with inner forms it is advantageous to consider the conjugation action of $\mathrm{SL}_n(\mathbb{C})$ on these two groups. It induces a natural action of $\mathrm{SL}_n(\mathbb{C})$ on the collection of Langlands parameters for $\mathrm{GL}_n(F)$. For any such parameter ϕ we can define

$$C(\phi) = Z_{\mathrm{SL}_n(\mathbb{C})}(\mathrm{im}\,\phi), \quad \text{and} \quad \mathcal{S}_\phi = C(\phi)/C(\phi)^\circ. \tag{3}$$

Notice that the centralizers are taken in $\mathrm{SL}_n(\mathbb{C})$ and not in the Langlands dual group.

Via the Langlands correspondence the nontrivial irreducible representations of $\mathcal{S}_\phi$ are associated with irreducible representations of non-split inner forms of $\mathrm{GL}_n(F)$. For example, consider a Langlands parameter ϕ for $\mathrm{GL}_2(F)$ which is elliptic, that is, whose image is not contained in any torus of $\mathrm{GL}_2(\mathbb{C})$. Then $\mathcal{S}_\phi = Z(\mathrm{SL}_2(\mathbb{C})) \cong \{\pm 1\}$. The pair $(\phi, \mathrm{triv}_{\mathcal{S}_\phi})$ parametrizes an essentially square-integrable representation of $\mathrm{GL}_2(F)$ and $(\phi, \mathrm{sgn}_{\mathcal{S}_\phi})$ parametrizes an irreducible representation of the inner form $D^\times$, where D denotes a noncommutative division algebra of dimension 4 over F.

The enhanced version of the local Langlands correspondence for all inner forms of general linear groups over non-archimedean local fields says:

Theorem 2.1 ([ABPS2, Theorem 1.1]). *There is a canonical bijection between:*

- *pairs (G, π) with $\pi \in Irr(G)$ and G an inner form of $GL_n(F)$, considered up to equivalence;*
- *$GL_n(\mathbb{C})$-conjugacy classes of pairs (ϕ, ρ) with $\phi \in \Phi(GL_n(F))$ and $\rho \in Irr(\mathcal{S}_\phi)$.*

Via the Kottwitz isomorphism [Kot, Proposition 6.4] every character of $Z(\mathrm{SL}_n(\mathbb{C}))$ determines a central simple F-algebra $\mathrm{M}_m(D)$. As $Z(\mathrm{SL}_n(\mathbb{C})) \subset C(\phi)$, for any Langlands parameter as above a character of $\mathcal{S}_\phi$ determines an inner form $\mathrm{GL}_m(D)$ of $\mathrm{GL}_n(F)$. In contrast with the usual LLC, our L-packets for inner forms of general linear groups need not be singletons. To be precise, the packet Π_ϕ contains the unique representation $\mathrm{rec}_{D,m}^{-1}(\phi)$ of $G = \mathrm{GL}_m(D)$ if ϕ is relevant for G, and no G-representations otherwise.

2.2 The Jacquet–Langlands Correspondence

A representation π of G is called essentially square-integrable if $\pi\big|_{G_{\mathrm{der}}}$ is square-integrable and there exists an unramified character χ of G such that $\pi \otimes \chi$ is unitary. We denote the set of (equivalence classes of) irreducible essentially square-integrable G-representations by $\mathrm{Irr}_{\mathrm{ess}L^2}(G)$. There is a natural bijection between $\mathrm{Irr}_{\mathrm{ess}L^2}(\mathrm{GL}_n(F))$ and $\mathrm{Irr}_{\mathrm{ess}L^2}(\mathrm{GL}_m(D))$, discovered first for $\mathrm{GL}_2(F)$ by Jacquet and Langlands [JaLa]. The local Langlands correspondence for $\mathrm{GL}_m(D)$ is constructed with the help of the Jacquet–Langlands correspondence. Here we recall some useful properties of the latter correspondence.

Theorem 2.2. *Let $GL_m(D)$ be an inner form of $GL_n(F)$. There exists a canonical bijection*

$$JL : Irr_{essL^2}(GL_n(F)) \to Irr_{essL^2}(GL_m(D))$$

with the following properties:

(*a*) *There is a canonical identification of the semisimple elliptic conjugacy classes in $GL_n(F)$ with those in $GL_m(D)$. Let $g \in GL_n(F)$ and $g' \in GL_m(D)$ be semisimple elliptic elements in the same conjugacy class and let θ_π be the character of $\pi \in Irr_{essL^2}(GL_n(F))$. Then*

$$(-1)^n \theta_\pi(g) = (-1)^m \theta_{JL(\pi)}(g').$$

(*b*) *JL preserves twists with characters of $F^\times$:*

$$JL(\pi \otimes \chi \circ \det) = JL(\pi) \otimes \chi \circ \mathrm{Nrd}.$$

(*c*) *JL respects contregredient: $JL(\pi^\vee) = JL(\pi)^\vee$.*

(*d*) *Let P' be a standard parabolic subgroup of $GL_m(D)$, with Levi factor $M' = \prod_i GL_{m_i}(D)$. Let P be the corresponding standard parabolic subgroup of $GL_n(F)$, with Levi factor $M = \prod_i GL_{dm_i}(F)$. Then the Jacquet modules $r_P^{GL_n(F)}(\pi)$ and $r_{P'}^{GL_m(D)}(JL(\pi))$ are either both zero or both irreducible and essentially square-integrable. In the latter case*

$$JL\big(r_P^{GL_n(F)}(\pi)\big) = r_{P'}^{GL_m(D)}(JL(\pi)).$$

In other words, JL and its version for M and M' respect Jacquet restriction.

(*e*) *JL preserves supercuspidality.*

(*f*) *$JL(St_{GL_n(F)}) = St_{GL_m(D)}$, where St_G denotes the Steinberg representation of G.*

(*g*) *JL preserves γ-factors:*

$$\gamma(s, JL(\pi), \psi) = \gamma(s, \pi, \psi) \quad \text{for any nontrivial character } \psi \text{ of } F.$$

(*h*) *JL preserves L-functions: $L(s, JL(\pi)) = L(s, \pi)$.*

(*i*) *JL preserves ϵ-factors: $\epsilon(s, JL(\pi), \psi) = \epsilon(s, \pi, \psi)$.*

Proof. The correspondence, which is in fact characterized by property (a), is proven over p-adic fields in [DKV] and over local fields of positive characteristic in [Bad]. Properties (b) and (c) are obvious in view of (a). The same goes for property (f) in the case $m = 1$, because then $\mathrm{St}_{\mathrm{GL}_m(D)}$ is just the trivial representation of $D^\times$. For (d) see [Bad, §5], in particular Proposition B. Obviously (d) implies (e). Property (f) for $m > 1$ follows from (f) for $m = 1$ and property (d). Property (g) was proven over local function fields in [Bad, p. 741], with an argument that also works over p-adic fields.

Properties (h) and (i) were claimed in [DKV], with the difference that the ϵ-factors of π and $\mathrm{JL}(\pi)$ are said to agree only up to a sign $(-1)^{n+m}$. This sign is due to a convention that does not agree with [GoJa], which we use for the definition of ϵ-factors. Unfortunately the argument for (h) and (i) given in [DKV, §B.1] is incorrect. Instead, we will establish (h) by direct calculation.

Let ν_D denote the unramified character $g' \mapsto \|\mathrm{Nrd}\, g'\|_F$ of $\mathrm{GL}_m(D)$. Consider any $\pi' \in \mathrm{Irr}_{\mathrm{ess}L^2}(\mathrm{GL}_m(D))$. By [DKV, §B.2] or [Tad, §2] there exist:

- integers a, b, s_σ such that $ab = m$ and s_σ divides ad;
- an irreducible supercuspidal representation σ of $\mathrm{GL}_a(D)$,

such that π' is a constituent of the parabolically induced representation

$$\Pi' := I_{\mathrm{GL}_a(D)^b}^{\mathrm{GL}_m(D)}\Big(\nu_D^{s_\sigma \frac{1-b}{2}}\sigma \otimes \nu_D^{s_\sigma \frac{3-b}{2}}\sigma \otimes \cdots \otimes \nu_D^{s_\sigma \frac{b-1}{2}}\sigma\Big). \tag{4}$$

By [Jac, Proposition 2.3] the L-function of (4) is the product of L-functions of the inducing representations:

$$L(s, \Pi') = \prod_{k=1}^{b} L(s, \nu_D^{s_\sigma(k-(1+b)/2)}\sigma) = \prod_{k=1}^{b} L(s + s_\sigma(k - (1+b)/2), \sigma). \tag{5}$$

By definition $L(s, \pi')^{-1}$ is a monic polynomial in q^{-s}, and by [Jac, 2.7.4] it is a factor of the monic polynomial $L(s, \Pi')^{-1}$. Now there are two cases to be distinguished, depending on whether σ is an unramified representation of $D^\times$ or not.

Case 1: $a = 1, b = m$ and σ is unramified.

There exists an unramified character χ of $F^\times$ such that $\sigma = \chi \circ \mathrm{Nrd}$. By [DKV, §B.2] or [Tad, §2] (4) only has an essentially square-integrable subquotient if $s_\sigma = d$. Then $\pi' \cong \mathrm{St}_{\mathrm{GL}_m(D)} \otimes \chi \circ \mathrm{Nrd}$, so $\mathrm{JL}^{-1}(\pi') \cong St_{\mathrm{GL}_n(F)} \otimes \chi \circ \det$. With property (f) this enables us to compute the γ-factor. Let ϖ_F be a uniformizer of F, $\mathfrak{o}_F$ the ring of integers and $\mathfrak{p}_F$ its maximal ideal. Assume that ψ is trivial on $\mathfrak{p}_F$ but not on $\mathfrak{o}_F$. Then

$$\gamma(s, \pi', \psi) = \gamma(s, \mathrm{St}_{\mathrm{GL}_n(F)} \otimes \chi \circ \det, \psi) = (-1)^n q^{n/2} \frac{1 - q^{-s+(1-n)/2} \chi(\varpi_F)}{1 - q^{-s+(1+n)/2} \chi(\varpi_F)}. \tag{6}$$

By [GoJa, Proposition 4.4], (5) becomes

$$\prod_{k=1}^{m} L(s + d(k - (1+m)/2), \chi \circ \mathrm{Nrd}) = \prod_{k=1}^{m} L(s + d(k - (1+m)/2) + (d-1)/2, \chi). \tag{7}$$

Now we apply [GoJa, Theorem 7.11.4]. It is stated only for $\mathrm{GL}_n(F)$, but the proof with zeros and poles of L-functions goes through because we know $\gamma(s, \pi', \psi)$. We find that for the L-function of π' we need only the factor $k = m$ of (7):

$$L(s, \pi', \psi) = L(s + (n-1)/2, \chi) = (1 - q^{-s+(1-n)/2} \chi(\varpi_F))^{-1}.$$

In particular the whole calculation works with $d = 1$, so

$$L(s, \mathrm{St}_{\mathrm{GL}_m(D)} \otimes \chi \circ \mathrm{Nrd}) = L(s, \mathrm{St}_{\mathrm{GL}_n(F)} \otimes \chi \circ \det) = L(s + (n-1)/2, \chi). \tag{8}$$

Case 2: all other σ.

Then [GoJa, Proposition 5.11] says that $L(s, \sigma \otimes \chi) = 1$ for every unramified character χ of $\mathrm{GL}_a(D)$. Hence $L(s, \Pi') = 1$ by (5). We observed above that $L(s, \pi')^{-1}$ is a factor of $L(s, \Pi')$, so $L(s, \pi') = 1$. Because JL is bijective, $\mathrm{JL}^{-1}(\pi')$ is not an unramified twist of the Steinberg representation, so $L(s, \mathrm{JL}^{-1}(\pi')) = 1$ as well. This proves property (h).

In view of the relation

$$\epsilon(s, \pi, \psi) = \gamma(s, \pi, \psi) L(s, \pi) L(1 - s, \pi^\vee), \tag{9}$$

(i) follows directly from (c), (g) and (h). □

We record a particular consequence of Eqs. (6), (8) and (9):

$$\epsilon(s, \mathrm{St}_{\mathrm{GL}_m(D)} \otimes \chi \circ \mathrm{Nrd}, \psi) = (-1)^{n-1} \epsilon(s, \chi, \psi) = (-1)^{n-1} q^{s-1/2} \chi(\varpi_F^{-1}) \tag{10}$$

for any character ψ of F which is trivial on $\mathfrak{p}_F$ but not on $\mathfrak{o}_F$.

2.3 Depth for Langlands Parameters

Let F_s be a separable closure of F and let $\mathrm{Gal}(F_s/F)$ be the absolute Galois group of F. We recall some properties of its ramification groups (with respect to the upper numbering), as defined in [Ser, Remark IV.3.1]:

- $\mathrm{Gal}(F_s/F)^{-1} = \mathrm{Gal}(F_s/F)$ and $\mathrm{Gal}(F_s/F)^0 = \mathbf{I}_F$, the inertia group.
- For every $l \in \mathbb{R}_{\geq 0}$, $\mathrm{Gal}(F_s/F)^l$ is a compact subgroup of $\mathbf{I}_F$. It consists of all $\gamma \in \mathrm{Gal}(F_s/F)$ which, for every finite Galois extension E of F contained in F_s, act trivially on the ring $\mathfrak{o}_E/\mathfrak{p}_E^{i(l,E)}$ (where $i(l,E) \in \mathbb{Z}_{\geq 0}$ can be found with [Ser, §IV.3]).
- $l \in \mathbb{R}_{\geq 0}$ is called a jump of the filtration if

$$\mathrm{Gal}(F_s/F)^{l+} := \bigcap_{l' > l} \mathrm{Gal}(F_s/F)^{l'}$$

  does not equal $\mathrm{Gal}(F_s/F)^l$. The set of jumps of the filtration is countably infinite and need not consist of integers.

Recall [Bor] that a Langlands parameter for $\mathrm{GL}_m(D)$ is a continuous homomorphism

$$\phi : \mathbf{W}_F \times \mathrm{SL}_2(\mathbb{C}) \to \mathrm{GL}_n(\mathbb{C})$$

such that:

- $\phi(\mathbf{W}_F)$ consists of semisimple elements;
- $\phi\big|_{\mathrm{SL}_2(\mathbb{C})} : \mathrm{SL}_2(\mathbb{C}) \to \mathrm{GL}_n(\mathbb{C})$ is a morphism of complex algebraic groups;
- ϕ is relevant for $\mathrm{GL}_m(D)$. This means that the conjugacy class of a Levi subgroup of $\mathrm{GL}_n(\mathbb{C})$ minimally containing $\mathrm{im}(\phi)$ should correspond to a conjugacy of class of Levi subgroups of $\mathrm{GL}_m(D)$.

We define the depth of such a Langlands parameter as

$$d(\phi) := \inf\{l \geq 0 \mid \mathrm{Gal}(F_s/F)^{l+} \subset \ker \phi\}.$$

We say that $\phi \in \Phi(\mathrm{GL}_n(F))$ is *elliptic* if its image is not contained in any proper Levi subgroup of $\mathrm{GL}_n(\mathbb{C})$.

Let ψ be a nontrivial character of F and let $c(\psi)$ be the largest integer c such that ψ is trivial on $\mathfrak{p}_F^{-c}$. The ϵ factor of ϕ (and ψ) was defined in [Tat]. It takes the form

$$\epsilon(s, \phi, \psi) = \epsilon(0, \phi, \psi) q^{-s(a(\phi)+nc(\psi))} \text{ with } \epsilon(0, \phi, \psi) \in \mathbb{C}^\times. \tag{11}$$

Here $a(\phi) \in \mathbb{Z}_{\geq 0}$ is the Artin conductor of ϕ (called $f(\phi)$ in [Ser, §VI.2]). To study $a(\phi)$ it is convenient to rewrite ϕ in terms of the Weil–Deligne group. For $\gamma \in \mathbf{W}_F$ we write $\|\gamma\| = q$ if γ induces the automorphism $x \mapsto x^q$ on the residue field of F_s.

Define

$$\phi_0(\gamma) = \phi(\gamma, 1)\phi\left(1, \begin{pmatrix} \|\gamma\|^{1/2} & 0 \\ 0 & \|\gamma\|^{-1/2} \end{pmatrix}\right), \tag{12}$$

so ϕ_0 is a representation of $\mathbf{W}_F$ which agrees with ϕ on $\mathbf{I}_F$. Define $N \in \mathfrak{gl}_n(\mathbb{C})$ as the nilpotent element $\log \phi\big(1, \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)\big)$. Then (ϕ_0, N) is the Weil–Deligne representation of $\mathbf{W}_F \ltimes \mathbb{C}$ corresponding to ϕ.

Denote the vector space $\mathbb{C}^n$ endowed with the representation ϕ by V, and write $V_N = \ker(N : V \to V)$. By definition [Tat, §4.1.6]

$$a(\phi) = a(\phi_0) + \dim(V^{\mathbf{I}_F}/V_N^{\mathbf{I}_F}), \tag{13}$$

$$\epsilon(s, \phi, \psi) = \epsilon(0, \phi_0, \psi) \det\Big(-\operatorname{Frob}\big|_{V^{\mathbf{I}_F}/V_N^{\mathbf{I}_F}}\Big) q^{-s(a(\phi)+nc(\psi))}, \tag{14}$$

where Frob denotes a geometric Frobenius element of $\mathbf{W}_F$.

Lemma 2.3. *For any elliptic $\phi \in \Phi(GL_n(F))$*

$$d(\phi) := \begin{cases} 0 & \text{if } \mathbf{I}_F \subset \ker(\phi), \\ \dfrac{a(\phi)}{n} - 1 & \text{otherwise}, \end{cases} \tag{15}$$

Proof. This was proved in [ABPS2, Lemma 4.4] under the additional assumption $SL_2(\mathbb{C}) \subset \ker \phi$. We will reduce to that special case.

Since ϕ is elliptic, it defines an irreducible n-dimensional representation V of $\mathbf{W}_F \times SL_2(\mathbb{C})$. Hence there are irreducible representations (ϕ_1, V_1) of $\mathbf{W}_F$ and (ϕ_2, V_2) of $SL_2(\mathbb{C})$ such that

$$(\phi, V) = (\phi_1, V_1) \otimes (\phi_2, V_2). \tag{16}$$

In particular $V^{\mathbf{I}_F} = V_1^{\mathbf{I}_F} \otimes V_2$. Suppose first that $V_1^{\mathbf{I}_F} = V_1$. Then $\mathbf{I}_F \subset \ker \phi$, so $d(\phi) = 0$ by definition. Now suppose $V_1^{\mathbf{I}_F} \neq V_1$. As (ϕ_1, V_1) is irreducible and $\mathbf{I}_F$ is normal in $\mathbf{W}_F$, we must have $V_1^{\mathbf{I}_F} = 0$. Hence $V^{\mathbf{I}_F} = 0$, which by (12) and (13) implies $a(\phi) = a(\phi_0)$. By [Ser, Corollary VI.2.1'] $a(\phi_0)$ is additive in V and depends only on

$$\phi_0\big|_{\mathbf{I}_F} = \phi\big|_{\mathbf{I}_F} = \phi_1\big|_{\mathbf{I}_F} \otimes \mathrm{id}_{V_2}.$$

Now it follows from (16) that

$$a(\phi) = a(\phi_1) \dim V_2 = n a(\phi_1) / \dim V_1. \tag{17}$$

As $\ker \phi_1 \supset \mathrm{SL}_2(\mathbb{C})$ we may apply [ABPS2, Lemma 4.4], which together with (17) gives

$$d(\phi_1) = \frac{a(\phi_1)}{\dim V_1} - 1 = \frac{a(\phi)}{n} - 1.$$

To conclude, we note that $d(\phi_1) = d(\phi)$ by (16). □

2.4 The Depth of Representations of $GL_m(D)$

Let $k_D = \mathfrak{o}_D/\mathfrak{p}_D$ be the residual field of D. Let $\mathfrak{A}$ be a hereditary $\mathfrak{o}_F$-order in $\mathrm{M}_m(D)$. The Jacobson radical of $\mathfrak{A}$ will be denoted by $\mathfrak{P}$. Let $r = e_D(\mathfrak{A})$ and $e = e_F(\mathfrak{A})$ denote the integers defined by $\mathfrak{p}_D\mathfrak{A} = \mathfrak{P}^r$ and $\mathfrak{p}_F\mathfrak{A} = \mathfrak{P}^e$, respectively. We have

$$e_F(\mathfrak{A}) = d\, e_D(\mathfrak{A}). \tag{18}$$

The normalizer in G of $\mathfrak{A}^\times$ will be denoted by

$$\mathfrak{K}(\mathfrak{A}) := \left\{g \in G \,:\, g^{-1}\mathfrak{A}^\times g = \mathfrak{A}^\times\right\}.$$

Define a sequence of compact open subgroups of $G = \mathrm{GL}_m(D)$ by

$$U^0(\mathfrak{A}) := \mathfrak{A}^\times, \quad \text{and} \quad U^j(\mathfrak{A}) := 1 + \mathfrak{P}^j,\ \ j \geq 1.$$

Then $\mathfrak{A}^\times$ is a parahoric subgroup of G and $U^1(\mathfrak{A})$ is its pro-unipotent radical. We define the *normalized level* of an irreducible representation π of G to be

$$d(\pi) := \min\{j/e_F(\mathfrak{A})\}, \tag{19}$$

where $(j, \mathfrak{A})$ ranges over all pairs consisting of an integer $j \geq 0$ and a hereditary $\mathfrak{o}_F$-order $\mathfrak{A}$ in $\mathrm{M}_m(D)$ such that π contains the trivial character of $U^{j+1}(\mathfrak{A})$.

Remark 2.4. When π is a representation of $\mathrm{GL}_n(F)$, our notion of normalized level coincides with that of [BuHe2, §12.6]. However when π is a representation of $D^\times$ (or more generally of $\mathrm{GL}_m(D)$), the normalized level of π as defined above is not equal to the level $\ell_D(\pi)$ defined in [BuHe2, §54.1] (resp. $\ell(\pi)$ defined by Broussous in [BaBr, Théorème A.1.2]): we have

$$d(\pi) = \frac{1}{d}\,\ell_D(\pi) \quad \text{(resp.} \quad d(\pi) = \frac{1}{d}\,\ell(\pi).$$

This reflects the fact that we have divided by $e_F(\mathfrak{A})$ instead of $e_D(\mathfrak{A})$ in Eq. (19).

The following proposition will allow to use both results that were written in the setting of the normalized level, as general results on the depth in the sense of Moy and Prasad.

Proposition 2.5. *The normalized level of* $\pi \in Irr(G)$ *equals its Moy–Prasad depth.*

Proof. Let us denote the Moy–Prasad depth of (π, V_π) by $d_{\mathrm{MP}}(\pi)$ for the duration of this proof. For any point x of the Bruhat–Tits building $\mathcal{B}(G)$ of G, consider the Moy–Prasad filtrations $P_{x,r}, P_{x,r+}$ $(r \in \mathbb{R}_{\geq 0})$ of the parahoric subgroup $P_x \subset G$ [MoPr1, §2]. We normalize these filtrations by using the valuation on F_s which maps $F^\times$ onto $\mathbb{Z}$. Then $d_{\mathrm{MP}}(\pi)$ is the minimal $r \in \mathbb{R}_{\geq 0}$ such that $V_\pi^{P_{x,r+}} \neq 0$ for some $x \in \mathcal{B}(G)$, see [MoPr2, §3.4].

Any hereditary $\mathfrak{o}_F$-order $\mathfrak{A}$ in $\mathrm{M}_m(D)$ is associated with a unique facet $\mathcal{F}(\mathfrak{A})$ of $\mathcal{B}(G)$. The filtration $\{U^j(\mathfrak{A}) \mid j \in \mathbb{Z}_{\geq 0}\}$ was compared with the Moy–Prasad groups for $x \in \mathcal{F}(\mathfrak{A})$ by Broussous and Lemaire. Let $x_{\mathfrak{A}}$ be the barycenter of $\mathcal{F}(\mathfrak{A})$. From [BrLe, Proposition 4.2 and Appendix A] and the definition of $e_F(\mathfrak{A})$ we see that

$$U^j(\mathfrak{A}) = P_{x_{\mathfrak{A}}, j/e_F(\mathfrak{A})} \text{ for all } j \in \mathbb{Z}_{\geq 0}.$$

Hence the definitions of the normalized level and the Moy–Prasad depth are almost equivalent, the only difference being that for $d_{\mathrm{MP}}(\pi)$ we must consider all points of $\mathcal{B}(G)$, whereas for $d(\pi)$ we may only use barycenters of facets of $\mathcal{B}(G)$. Thus it remains to check the following claim: there exists a facet $\mathcal{F}$ of $\mathcal{B}(G)$ with barycenter $x_{\mathcal{F}}$, such that V_π has nonzero $P_{x_{\mathcal{F}}, d_{\mathrm{MP}}(\pi)+}$-invariant vectors.

This is easy to see with the explicit constructions of the groups $P_{x,r}$ at hand, but we prefer not to delve into those details here. In fact, since every chamber of $\mathcal{B}(G)$ intersects every G-orbit in $\mathcal{B}(G)$, it suffices to consider facets contained in the closure of a fixed "standard" chamber. Then the claim becomes equivalent to saying that $x_{\mathcal{F}}$ is an "optimal point" in the sense of [MoPr1, §6.1]. That is assured by [MoPr1, Remark 6.1], which is applicable because the root system of G is of type A_{m-1}. □

2.5 Conductors of Representations of $GL_m(D)$

Let $\epsilon(s, \pi, \psi)$ denote the Godement–Jacquet local constant [GoJa]. It takes the form

$$\epsilon(s, \pi, \psi) = \epsilon(0, \pi, \psi)\, q^{-f(\pi,\psi)s}, \quad \text{where } \epsilon(0, \pi, \psi) \in \mathbb{C}^\times. \tag{20}$$

Recall that $c(\psi)$ is the largest integer c such that $\mathfrak{p}_F^{-c} \subset \ker \psi$. In the previous section we had $c(\psi) = -1$.

A representation of $D^\times$ is called *unramified* if it is trivial on $\mathfrak{o}_D^\times$. An unramified representation of $D^\times$ is a character and has depth zero.

Proposition 2.6. *Let π be a supercuspidal irreducible representation of G. We have*

$$f(\pi,\psi) = \begin{cases} n\,(c(\psi)+1)-1 & \text{if } m=1 \text{ and } \pi \text{ is unramified,} \\ n\,\left(d(\pi)+1+c(\psi)\right) & \text{otherwise.} \end{cases} \tag{21}$$

Proof. We suppose first that $m = 1$ (so $d = n$) and π is unramified. The required formula can be read off from (10) if $c(\psi) = -1$. For general ψ, applying [BuFr1, Theorem 3.2.11] and taking into account [BuFr1, (1.2.7), (1.2.8), (1.2.10)], we obtain

$$f(\pi,\psi) = (d(1-d-dc(\psi))\cdot\left(-\frac{1}{d}\right) = d + dc(\psi) - 1.$$

Hence the first case of Eq. (21) holds.

From now on, we will assume that $m \geq 2$ or π is ramified. Then by combining [BaBr, Théorème A.2.1] with the fact that the Godement–Jacquet L-function $L(s,\pi)$ is 1, we see that π satisfies the conditions of Theorem 3.3.8 of [BuFr2]. Choose $(j,\mathfrak{A})$ as in (19), so $d(\pi) = j/e_F(\mathfrak{A})$ and π contains the trivial representation of $U^{j+1}(\mathfrak{A})$. Recall that $n = md$. By applying the formula of [BuFr2, Theorem 3.3.8 (iv)], we obtain

$$q^{f(\pi,\psi)} = \left[\mathfrak{A} : \mathfrak{p}_F^{c(\psi)+1}\mathfrak{P}^j\right]^{1/n}.$$

On the other hand, the $\mathfrak{o}_F$-order $\mathfrak{A}$ is G-conjugate to the standard principal $\mathfrak{o}_F$-order of $\mathrm{M}_m(D)$ defined by the partition $(t,\ldots,t)$ (r-times) of m, where $m = rt$ and $r = e_D(\mathfrak{A})$. Hence we have $\mathfrak{A}/\mathfrak{P} \simeq (\mathrm{M}_t(k_D))^r$. It follows that

$$[\mathfrak{A} : \mathfrak{P}] = (q^d)^{t^2 r} = q^{drt^2}.$$

Hence we get

$$f(\pi,\psi) = \frac{drt^2(j+e+ec(\psi))}{n} = n\left(\frac{j}{e}+1+c(\psi)\right),$$

since $drt^2 = nt = n^2/e$.

On the other hand, it follows from [SéSt1, Corollaire 5.22] that there exists a maximal simple type (J,λ) in G, and an extension Λ of λ to the normalizer $\bar{J} = \mathrm{N}_G(\lambda)$ of λ, such that

$$\pi = \mathrm{c}\!-\!\mathrm{Ind}_{\bar{J}}^G\Lambda.$$

By the construction of the type (J, λ), we have $d(\pi) \leq j/e$. Conversely, let $[\mathfrak{A}', j', j' - 1, \beta']$ be a stratum contained in π. Then if $[\mathfrak{A}', j', j' - 1, \beta']$ is such that its normalized level j'/e' is minimal among the normalized levels of all the strata contained in π, it is necessarily fundamental [Bro, Theorem 1.2.1. (ii)]. Since all the fundamental strata contained in π have the same normalized level [BaBr, Théorème A.1.2], we get $j/e = d(\pi)$. ☐

Theorem 2.7 below proves the validity of Conjecture 4.3 of [LaRa]. In the case when F has characteristic 0, it is due to Lansky and Raghuram for the groups $\mathrm{GL}_n(F)$ and $D^\times$, [LaRa, Theorem 3.1], and for certain representations of $\mathrm{GL}_2(D)$, [LaRa, Theorem 4.1]. Our proof is inspired by those of these results.

Theorem 2.7. *The depth $d(\pi)$ and the conductor $f(\pi) := f(\pi, \psi) - nc(\psi)$ of each essentially square-integrable irreducible representation π of $GL_m(D)$ are linked by the following relation:*

$$d(\pi) = \begin{cases} 0 & \text{if } \pi \text{ is an unramified twist of } St_{GL_m(D)}, \\ \dfrac{f(\pi) - n}{n} & \text{otherwise.} \end{cases} \tag{22}$$

In particular

$$d(\pi) = \max\left\{\frac{f(\pi) - n}{n}, 0\right\}. \tag{23}$$

Proof. Let $\pi \in \mathrm{Irr}_{\mathrm{ess}L^2}(\mathrm{GL}_m(D))$. We use the same notation as for π' in the proof of Theorem 2.2.h, so π is constituent of

$$I_{\mathrm{GL}_a(D)^b}^{\mathrm{GL}_m(D)}\left(\nu_D^{s_\sigma\frac{(1-b)}{2}}\sigma \otimes \nu_D^{s_\sigma\frac{(3-b)}{2}}\sigma \otimes \cdots \otimes \nu_D^{s_\sigma\frac{(b-1)}{2}}\sigma\right),$$

where $\sigma \in \mathrm{Irr}(\mathrm{GL}_a(D))$ is supercuspidal. Since the depth is preserved by parabolic induction [MoPr2, Theorem 5.2], we get

$$d(\pi) = d\left(\nu_D^{s_\sigma\frac{(1-b)}{2}}\sigma \otimes \nu_D^{s_\sigma\frac{(3-b)}{2}}\sigma \otimes \cdots \otimes \nu_D^{s_\sigma\frac{(b-1)}{2}}\sigma\right).$$

It follows that

$$d(\pi) = d(\sigma). \tag{24}$$

We will apply Proposition 2.6 to the supercuspidal representation σ of $\mathrm{GL}_a(D)$. In the special case σ is an unramified representation of $D^\times$ (hence $a = 1$ in this case), Eq. (21) gives

$$f(\sigma, \psi) = d\,(c(\psi) + 1) - 1,$$

that is, $f(\sigma) = d - 1$. Hence we get

$$\frac{f(\sigma) - d}{d} = -\frac{1}{d}.$$

Then it implies that

$$\max\left\{\frac{f(\sigma) - d}{d}, 0\right\} = \max\left\{-\frac{1}{d}, 0\right\} = 0 = d(\sigma),$$

in other words, Eq. (23) holds for the unramified representations of $D^{\times}$.

In the other cases (that is, $a \neq 1$ or σ is ramified), (21) gives $f(\sigma) = ad(d(\sigma)+1)$, that is,

$$\frac{f(\sigma)}{ad} = d(\sigma) + 1. \tag{25}$$

Since $d(\sigma) \geq 0$ (by definition of the depth), we obtain that

$$d(\sigma) = \max\left\{\frac{f(\sigma) - ad}{ad}, 0\right\}. \tag{26}$$

Hence (22) holds for every supercuspidal irreducible representation of $\mathrm{GL}_a(D)$, with $a \geq 1$ an arbitrary integer.

Recall that s_σ is an integer dividing ad, say $ad = a^* s_\sigma$ with $a^* \in \mathbb{Z}$. The image $\mathrm{JL}^{-1}(\sigma)$ of σ under the Jacquet–Langlands correspondence is equivalent to the Langlands quotient of the parabolically induced representation

$$I_{\mathrm{GL}_{a^*}(F)^{s_\sigma}}^{\mathrm{GL}_{a^* s_\sigma}(F)}\left(\nu_F^{\frac{(1-s_\sigma)}{2}}\sigma^* \otimes \nu_F^{\frac{(3-s_\sigma)}{2}}\sigma^* \otimes \cdots \otimes \nu_F^{\frac{(s_\sigma-1)}{2}}\sigma^*\right),$$

where σ^* is a unitary supercuspidal irreducible representation of $\mathrm{GL}_{a^*}(F)$ and $\nu_F(g^*) = |\det(g^*)|_F$.

The representation $\mathrm{JL}^{-1}(\pi)$ is equivalent to a constituent of the parabolically induced representation

$$I_{\mathrm{GL}_{a^*}(F)^{bs_\sigma}}^{\mathrm{GL}_{adb}(F)}\left(\nu_F^{\frac{(1-bs_\sigma)}{2}}\sigma^* \otimes \nu_F^{\frac{(3-bs_\sigma)}{2}}\sigma^* \otimes \cdots \otimes \nu_F^{\frac{(bs_\sigma-1)}{2}}\sigma^*\right).$$

We recall from [Hen, §2.6] the formula describing the epsilon factor of $\mathrm{JL}^{-1}(\pi)$ in terms of the local factors of σ^*:

$$\epsilon(s, \mathrm{JL}^{-1}(\pi), \psi) = \prod_{i=0}^{bs_\sigma-1} \epsilon(s+i, \sigma^*, \psi) \prod_{j=0}^{bs_\sigma-2} \frac{L(-s-j, \check{\sigma}^*)}{L(s+j, \sigma^*)}. \tag{27}$$

Since the Jacquet–Langlands correspondence preserves the ϵ-factors (see Theorem 2.2 i) we have

$$\epsilon(s, \mathrm{JL}^{-1}(\pi), \psi) = \epsilon(s, \pi, \psi).$$

Thus we have obtained the following formula:

$$\epsilon(s, \pi, \psi) = \prod_{i=0}^{bs_\sigma - 1} \epsilon(s + i, \sigma^*, \psi) \prod_{j=0}^{bs_\sigma - 2} \frac{L(-s - j, \check{\sigma}^*)}{L(s + j, \sigma^*)}. \tag{28}$$

If $\pi = \mathrm{St}_{\mathrm{GL}_m(D)} \otimes \chi$ for some unramified character χ of $D^\times$, it follows from (10) that $f(\pi, \psi) = -1$ in the case where $c(\psi) = -1$, hence we obtain

$$f(\pi) = n - 1. \tag{29}$$

From now on we assume π is not equivalent to a representation of the form $\mathrm{St}_{\mathrm{GL}_m(D)} \otimes \chi$, with χ an unramified character of $D^\times$ (that is, we have $m \neq 1$ or σ ramified). Then Theorem 2.2 b and f implies that similarly $\mathrm{JL}^{-1}(\pi)$ is not a twist of $\mathrm{St}_{\mathrm{GL}_n(F)}$ by an unramified character of $F^\times$. Thus we have $a^* \neq 1$ or σ^* ramified. It follows that $L(-s - j, \check{\sigma}^*) = L(s + j, \sigma^*) = 1$, and we obtain from (28) that

$$f(\pi) = bs_\sigma f(\sigma^*). \tag{30}$$

In the special case when $b = 1$ Eq. (30) gives

$$f(\sigma) = s_\sigma f(\sigma^*). \tag{31}$$

Then using (24) and (26) we get

$$d(\pi) = d(\sigma) = \max\left\{\frac{bs_\sigma f(\sigma^*) - \mathrm{bad}}{\mathrm{bad}}, 0\right\} = \max\left\{\frac{f(\pi) - n}{n}, 0\right\}. \square \tag{32}$$

2.6 Depth Preservation

Corollary 2.8. *The Jacquet–Langlands correspondence preserves the depth of essentially square-integrable representations of* $GL_m(D)$.

Proof. Theorem 2.2.i shows in particular that the Jacquet–Langlands correspondence preserves conductors. Now Theorem 2.7 shows that it preserves depths as well. □

Theorems 2.7 and 2.2 are also the crucial steps to show that the local Langlands correspondence for inner forms of $\mathrm{GL}_m(D)$ preserves depths. With similar considerations we show that it also preserves L-functions, ϵ-factors and γ-factors. We abbreviate these three to "local factors". For the basic properties of the local factors of Langlands parameters we refer to [Tat].

Theorem 2.9. *The local Langlands correspondence for representations of $GL_m(D)$ preserves L-functions, ϵ-factors, γ-factors and depths. In other words, for every irreducible smooth representation π of $GL_m(D)$:*

$$\begin{aligned} L(s,\pi) &= L(s, \mathrm{rec}_{D,m}(\pi)), \\ \epsilon(s,\pi,\psi) &= \epsilon(s, \mathrm{rec}_{D,m}(\pi), \psi), \\ \gamma(s,\pi,\psi) &= \gamma(s, \mathrm{rec}_{D,m}(\pi), \psi) \\ d(\pi) &= d(\mathrm{rec}_{D,m}(\pi)). \end{aligned}$$

Proof. It is well known that the local Langlands correspondence for $\mathrm{GL}_n(F)$ preserves local factors, see the introduction of [HaTa].

Assume first that π is essentially square-integrable. Recall the notations of the ϵ factors of π and of $\phi := \mathrm{rec}_{D,m}(\pi) \in \Phi(\mathrm{GL}_m(D))$ from (11) and (20). By definition

$$\mathrm{rec}_{D,m}(\pi) = \mathrm{rec}_{F,n}(\mathrm{JL}^{-1}(\pi)),$$

so by Theorem 2.2 $\mathrm{rec}_{D,m}$ preserves the ϵ-factors of π:

$$\epsilon(0,\phi,\psi)\, q^{-s(a(\phi)+nc(\psi))} = \epsilon(s,\phi,\psi) = \epsilon(s,\pi,\psi) = \epsilon(0,\pi,\psi) q^{-sf(\pi,\psi)}. \quad (33)$$

Hence, with the notation from Theorem 2.7:

$$f(\pi) = f(\pi,\psi) - nc(\psi) = a(\phi). \quad (34)$$

The properties of $\mathrm{rec}_{F,n}$ imply that ϕ is elliptic. By combining Lemma 2.3 with Theorem 2.7 and (34), we obtain that $d(\phi) = d(\pi)$ whenever π is essentially square-integrable.

Now let π be any irreducible representation of $\mathrm{GL}_m(D)$. By the Langlands classification, there exist a parabolic subgroup $P \subset \mathrm{GL}_m(D)$ with Levi factor M and an irreducible essentially square-integrable representation ω of M, such that π is a quotient of $I_P^{\mathrm{GL}_m(D)}(\omega)$. Moy and Prasad proved in [MoPr2, Theorem 5.2] that π and ω have the same depth. By [Jac, Theorem 3.4] π and ω have the same L-functions and ϵ-factors and by [Jac, (2.3) and (2.7.3)] they also have the same γ-factors.

On the other hand, M is isomorphic to a product of groups of the form $\mathrm{GL}_{m_i}(D)$, so the local Langlands correspondence for M is simply the product of that for the $\mathrm{GL}_{m_i}(D)$. The Langlands parameters $\mathrm{rec}_{D,m}(\pi)$ and $\mathrm{rec}_M(\omega)$ are related via an inclusion of the complex dual groups $\prod_i \mathrm{GL}_{dm_i}(\mathbb{C}) \to \mathrm{GL}_n(\mathbb{C})$. Hence these two Langlands parameters also have the same depth and local factors.

Because we already proved that the LLC preserves depths for essentially square-integrable representations of $\mathrm{GL}_m(D)$ or M, we can conclude that

$$d(\pi) = d(\omega) = d(\mathrm{rec}_M(\omega)) = d(\mathrm{rec}_{D,m}(\pi)),$$

and similarly for the local factors. □

3 The Local Langlands Correspondence for Inner Forms of $\mathrm{SL}_n(F)$

3.1 The Statement of the Correspondence

Recall that F is a non-archimedean local field and that the equivalence classes of inner forms of $\mathrm{SL}_n(F)$ are in bijection with the isomorphism classes of central simple F-algebras of dimension n^2, via $\mathrm{M}_m(D) \mapsto \mathrm{GL}_m(D)_{\mathrm{der}}$. As mentioned after Theorem 2.1, every character of $Z(\mathrm{SL}_n(\mathbb{C}))$ gives rise to such an algebra via the Kottwitz isomorphism.

The local Langlands correspondence for $\mathrm{GL}_m(D)_{\mathrm{der}}$ is implied by that for $\mathrm{GL}_m(D)$, in the following way. A Langlands parameter

$$\phi : \mathbf{W}_F \times \mathrm{SL}_2(\mathbb{C}) \to \mathrm{PGL}_n(\mathbb{C})$$

which is relevant for $\mathrm{GL}_m(D)_{\mathrm{der}}$ can be lifted it to a Langlands parameter

$$\overline{\phi} : \mathbf{W}_F \times \mathrm{SL}_2(\mathbb{C}) \to \mathrm{GL}_n(\mathbb{C})$$

which is relevant for $\mathrm{GL}_m(D)$, by Weil [Wei]. Then $\mathrm{rec}_{m,D}^{-1}(\overline{\phi})$ is an irreducible representation of $\mathrm{GL}_m(D)$ which, upon restriction to $\mathrm{GL}_m(D)_{\mathrm{der}}$, decomposes as a finite direct sum of irreducible representations. The packet $\Pi_\phi(\mathrm{GL}_m(D)_{\mathrm{der}})$ is defined as the set of irreducible constituents of $\mathrm{Res}^{\mathrm{GL}_m(D)}_{\mathrm{GL}_m(D)_{\mathrm{der}}} \mathrm{rec}_{m,D}^{-1}(\overline{\phi})$.

For these groups it is more interesting to consider the enhanced Langlands correspondence, where ϕ is supplemented with an irreducible representation of a finite group. In addition to the groups defined in (3), we write

$$\mathcal{Z}_\phi = Z(\mathrm{SL}_n(\mathbb{C}))/Z(\mathrm{SL}_n(\mathbb{C})) \cap C(\phi)^\circ \cong Z(\mathrm{SL}_n(\mathbb{C}))C(\phi)^\circ/C(\phi)^\circ. \tag{35}$$

Any character of $\mathcal{Z}_\phi$ determines a character of $Z(\mathrm{SL}_n(\mathbb{C}))$, and hence an inner form of $\mathrm{SL}_n(F)$. An enhanced Langlands parameter is a pair (ϕ, ρ) with $\rho \in \mathrm{Irr}(\mathcal{S}_\phi)$. The groups in (3), (35) are related to the more usual component group

$$S_\phi := Z_{\mathrm{PGL}_n(\mathbb{C})}(\mathrm{im}\,\phi)/Z_{\mathrm{PGL}_n(\mathbb{C})}(\mathrm{im}\,\phi)^\circ$$

by the short exact sequence

$$1 \to \mathcal{Z}_\phi \to \mathcal{S}_\phi \to S_\phi \to 1.$$

Hence $\mathcal{S}_\phi$ has more irreducible representations than S_ϕ. Via the enhanced Langlands correspondence the additional ones are associated with irreducible representations of non-split inner forms of $\mathrm{SL}_n(F)$. The following result is due to Hiraga and Saito [HiSa, Theorem 12.7] for generic representations of $\mathrm{GL}_m(D)$ when char $F = 0$.

Theorem 3.1 ([ABPS2, Theorem 1.2]). *There exists a bijective correspondence between:*

- *pairs $(GL_m(D)_{der}, \pi)$ with $\pi \in Irr(GL_m(D)_{der})$ and $GL_m(D)_{der}$ an inner form of $SL_n(F)$, considered up to equivalence;*
- *$SL_n(\mathbb{C})$-conjugacy classes of pairs (ϕ, ρ) with $\phi \in \Phi(SL_n(F))$ and $\rho \in Irr(\mathcal{S}_\phi)$.*

Here the group $GL_m(D)_{der}$ determines $\rho\big|_{\mathcal{Z}_\phi}$ and conversely. The correspondence satisfies the desired properties from [Bor, §10.3], with respect to restriction from inner forms of $GL_n(F)$, temperedness and essential square-integrability of representations.

We remark that the above bijection need not be canonical if $\Pi_\phi(\mathrm{GL}_m(D)_{\mathrm{der}})$ has more than one element.

3.2 The Depth of Representations of $GL_m(D)_{der}$

For the depth of an irreducible representation of $\mathrm{GL}_m(D)_{\mathrm{der}}$ there are two candidates. Besides the Moy–Prasad depth one can define the normalized level, just as in (19). This was done for representations of $\mathrm{SL}_n(F)$ in [BuKu]. However, Proposition 2.5 quickly reveals that these two notions agree:

Proposition 3.2. *The Moy–Prasad depth of an irreducible representation of $GL_m(D)_{der}$ equals its normalized level.*

Proof. Let us compare the descriptions of the two kinds of depth with those given in the proof of Proposition 2.5. By definition $\mathrm{GL}_m(D)$ and $\mathrm{GL}_m(D)_{\mathrm{der}}$ have the same Bruhat–Tits building. The Moy–Prasad depth is defined in terms of the groups

$$P'_{x,r} = P_{x,r} \cap \mathrm{GL}_m(D)_{\mathrm{der}} \quad \text{with} \quad x \in \mathcal{B}(\mathrm{GL}_m(D)). \tag{36}$$

The normalized level is expressed with the groups

$$U^j(\mathfrak{A})' = U^j(\mathfrak{A}) \cap \mathrm{GL}_m(D)_{\mathrm{der}},$$

where $\mathfrak{A}$ is a hereditary $\mathfrak{o}_F$-order in $M_m(D)$. With these groups instead of $P_{x,r}$ and $U^j(\mathfrak{A})$ the entire proof of Proposition 2.5 carries over to $\mathrm{GL}_m(D)_{\mathrm{der}}$. □

It turns out that the depth of an irreducible $\mathrm{GL}_m(D)_{\mathrm{der}}$-representation π behaves nicely with respect to restriction from $\mathrm{GL}_m(D)$. To be precise, equals the minimum of the depths of the irreducible $\mathrm{GL}_m(D)$-representations that contain π. (Notice that this minimum is always attained because all depths for inner forms of $\mathrm{GL}_n(F)$ lie in $\frac{1}{n}\mathbb{Z}$.)

Proposition 3.3. *Let $\pi \in Irr(GL_m(D)_{der})$ and let $\overline{\pi} \in Irr(GL_m(D))$ be such that*

- *π is a direct summand of* $\mathrm{Res}^{GL_m(D)}_{GL_m(D)_{der}}(\overline{\pi})$*;*
- *$d(\overline{\pi}) \leq d(\overline{\pi} \otimes \chi \circ Nrd)$ for every character χ of $F^\times$.*

Then $d(\pi) = d(\overline{\pi})$.

Proof. In the case $G = \mathrm{GL}_n(F)$, this is guaranteed by Proposition 3.2 and [BuKu, Proposition 1.7.iii]. The same proof works for $\mathrm{GL}_m(D)$ but this would be cumbersome, one would have to check that everything in [BuKu, pp. 265–268] also works with a division algebra instead of a field.

Instead, we select some parts of [BuKu, §1] to provide a shorter proof. Pick a $x \in \mathcal{B}(G)$ such that $(\overline{\pi}, V)$ has nonzero vector fixed by $P_{x,d(\overline{\pi})+}$. Then

$$V^{P'_{x,d(\overline{\pi})+}} \supset V^{P_{x,d(\overline{\pi})+}} \neq 0,$$

so there is an irreducible $\mathrm{GL}_m(D)_{\mathrm{der}}$-subrepresentation (π_1, V_1) of $\overline{\pi}$ with

$$V_1^{P'_{x,d(\overline{\pi})+}} \neq 0 \quad \text{and} \quad d(\pi_1) \leq d(\overline{\pi}).$$

Since $\overline{\pi}$ is irreducible, π_1 is isomorphic to a $\mathrm{GL}_m(D)$-conjugate of π. Conjugation by $g \in \mathrm{GL}_m(D)$ sends any Moy–Prasad group $P_{y,r}$ to $P_{g(y),r}$. So this operation preserves depths and

$$d(\pi) = d(\pi_1) \leq d(\overline{\pi}). \tag{37}$$

Suppose now that $d(\pi) < d(\overline{\pi})$. Take a nonzero $v \in V^{P'_{x,d(\pi)+}}$ and consider

$$V_v := \mathrm{span}\{\overline{\pi}(g)v \mid g \in P_{x,d(\pi)+}\}.$$

As $P'_{x,d(\pi)+}$ is normal in $P_{x,d(\pi)+}$, it acts trivially on V_v, and V_v can be regarded as representation of

$$P_{x,d(\pi)+}/P'_{x,d(\pi)+} \cong \mathrm{Nrd}(P_{x,d(\pi)+}) \subset F^\times.$$

Hence there is a character χ of $F^\times$ such that $\chi^{-1} \circ \mathrm{Nrd}$ appears in the action of $P_{x,d(\pi)+}$ on V_v. Then the irreducible $\mathrm{GL}_m(D)$-representation $\overline{\pi} \otimes \chi \circ \mathrm{Nrd}$ has a nonzero vector fixed by $P_{x,d(\pi)+}$, so

$$d(\overline{\pi} \otimes \chi \circ \mathrm{Nrd}) \leq d(\pi) < d(\overline{\pi}).$$

This contradicts the assumptions of proposition, so (37) must be an equality. □

3.3 The Depth of Langlands Parameters for $GL_m(D)_{der}$

The depth of a Langlands parameter $\phi : \mathbf{W}_F \times \mathrm{SL}_2(\mathbb{C}) \to \mathrm{PGL}_n(\mathbb{C})$ for an inner form of $\mathrm{SL}_n(F)$ is defined as in Sect. 2.3:

$$d(\phi) = \inf\{l \in \mathbb{R}_{\geq 0} \mid \mathrm{Gal}(F_s/F)^{l+} \subset \ker \phi\}.$$

The following result may be considered as the non-archimedean analogue of [ChKa, Theorem 1] in the case of the geometric local Langlands correspondence.

Corollary 3.4. *Let $\pi \in Irr(GL_m(D)_{der})$ with Langlands parameter $\phi \in \Phi(SL_n(F))$. Then $d(\pi) \geq d(\phi)$.*

Proof. Let $\overline{\pi}$ be as in Proposition 3.3, so $d(\overline{\pi}) = d(\pi)$. Put $\overline{\phi} = \mathrm{rec}_{D,m}(\overline{\pi})$, this is a lift of ϕ to $\mathrm{GL}_n(\mathbb{C})$ and Theorem 2.9 says that $d(\overline{\phi}) = d(\overline{\pi})$.

We remark that by the compatibility of the LLC with character twists

$$d(\overline{\phi}) \leq d(\overline{\phi} \otimes \gamma) \text{ for every character } \gamma \text{ of } \mathbf{W}_F. \tag{38}$$

For any lift $\overline{\phi} \in \Phi(\mathrm{GL}_n(F))$ of ϕ we have $\ker \overline{\phi} \subset \ker \phi$, so $d(\overline{\phi}) \geq d(\phi)$. □

It is possible that the inequality in Corollary 3.4 is strict. The following example was pointed out to the authors by Mark Reeder.

Example 3.5. Take $F = \mathbb{Q}_2$ and a Langlands parameter $\phi : \mathbf{W}_{\mathbb{Q}_2} \to \mathrm{PGL}_2(\mathbb{C})$ which is trivial on $\mathrm{SL}_2(\mathbb{C})$ and has image isomorphic to the symmetric group S_4. (Such a L-parameter exists, see, for example, [Wei].) We claim that $d(\phi) = 1/3$.

Let Ad denote the adjoint representation of $\mathrm{PGL}_2(\mathbb{C})$ on $\mathfrak{sl}_2(\mathbb{C}) = \mathrm{Lie}(\mathrm{PGL}_2(\mathbb{C}))$. Then $\mathrm{Ad} \circ \phi$ is an irreducible three-dimensional representation of $\mathbf{W}_{\mathbb{Q}_2}$. Since $\mathrm{PGL}_2(\mathbb{C})$ is the adjoint group of $\mathfrak{sl}_2(\mathbb{C})$, $\mathrm{Ad} \circ \phi$ has the same kernel and hence the same depth as ϕ. One can check that $\mathrm{Ad}(\phi(\mathbf{I}_F)) \cong A_4$ and that the image of the wild inertia subgroup $\mathbf{P}_F$ is isomorphic to the Klein four group. With the formula [GrRe, (1)] for the Artin conductor we find that $a(\mathrm{Ad} \circ \phi) = 4$. By Lemma 2.3 (with $n = 3$) $d(\mathrm{Ad} \circ \phi) = 1/3$.

Let $\overline{\phi} : \mathbf{W}_{\mathbb{Q}_2} \to \mathrm{GL}_2(\mathbb{C})$ be a lift of ϕ. This is an irreducible two-dimensional representation. We claim that $d(\overline{\phi}) \geq 1/2$.

With a suitable basis transformation we can achieve that

$$\phi(\mathbf{P}_{\mathbb{Q}_2}) = \{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & i \\ i & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} -i & 0 \\ 0 & i \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)\} \subset \mathrm{PGL}_2(\mathbb{C}).$$

Let $1, w_2, w_3, w_4 \in \mathbf{P}_{\mathbb{Q}_2}$ be preimages of these elements. Irrespective of the choice of the lift of ϕ we have

$$[\overline{\phi}(w_3), \overline{\phi}(w_4)] = [\left(\begin{smallmatrix} -i & 0 \\ 0 & i \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)] = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{C}).$$

Now we see the problem: since $\phi|_{\mathbf{P}_{\mathbb{Q}_2}}$ is not a sum of characters, commutators pop up in $\overline{\phi}|_{\mathbf{P}_{\mathbb{Q}_2}}$, and these will cause more ramification.

Put $E = F_s^{\ker \overline{\phi}}$ and endow $\mathrm{Gal}(E/\mathbb{Q}_2) \cong \overline{\phi}(\mathbf{W}_{\mathbb{Q}_2})$ with the lower numbered filtration. The image of $\mathbf{P}_{\mathbb{Q}_2}$ is $\mathrm{Gal}(E/\mathbb{Q}_2)_1$ and $[w_3, w_4] \in [\mathrm{Gal}(E/\mathbb{Q}_2)_1, \mathrm{Gal}(E/\mathbb{Q}_2)_1]$. Then $[w_3, w_4] \in \mathrm{Gal}(E/\mathbb{Q}_2)_3$ by [Ser, Proposition IV.2.10], so $\overline{\phi}$ is nontrivial on this ramification group. If we lift ϕ with as little ramification as possible, $\overline{\phi}(\mathbf{W}_{\mathbb{Q}_2})$ is an index 2 central extension of S_4. Writing $d_j = |\overline{\phi}(\mathrm{Gal}(E/\mathbb{Q}_2)_j)|$, we have

$$d_0 = 24, d_1 = 8, d_2 = d_3 = 2 \text{ and } d_j = 1 \text{ for } j > 3.$$

The formula [GrRe, (1)] gives

$$a(\overline{\phi}) = \frac{\dim(\overline{\phi})}{d_0} \sum_{j \geq 0 : d_j > 1} d_j = \frac{2}{24}(24 + 8 + 2 + 2) = 3.$$

Now Lemma 2.3 says that $d(\overline{\phi}) = 1/2$.

To show that Corollary 3.4 is in many cases an equality, we will make use of several well-known properties of the Artin reciprocity map $\mathbf{a}_F : \mathbf{W}_F \to F^\times$. In particular:

Theorem 3.6. $\mathbf{a}_F(\mathrm{Gal}(F_s/F)^l) = U_F^{\lceil l \rceil}$ *for all* $l \in \mathbb{R}_{\geq 0}$.

Proof. For any finite abelian extension E/F, [Ser, Corollary 3 to Theorem XV.2.1] says that the Artin reciprocity map gives an isomorphism

$$\mathbf{a}_F \colon \mathrm{Gal}(E/F)^l \to U_F^{\lceil l \rceil} / \big(\mathrm{N}_{E/F}(E^\times) \cap U_F^{\lceil l \rceil}\big). \tag{39}$$

Let F_s^{ab} be the maximal abelian extension of F contained in F_s. Taking the projective limit of (39) over all finite subextensions of F_s^{ab}/F, we obtain an isomorphism

$$\mathbf{a}_F : \mathrm{Gal}(F_s^{\mathrm{ab}}/F)^l \to U_F^{\lceil l \rceil}.$$

We note that $\mathrm{Gal}(F_s^{\mathrm{ab}}/F)$ is the quotient of $\mathrm{Gal}(F_s/F)$ modulo the closure of its commutator subgroup. Hence $\mathbf{a}_F : \mathbf{W}_F \to F^\times$ factors via $\mathrm{Gal}(F_s^{\mathrm{ab}}/F)$. □

Recall from [BuHe1] that a Langlands parameter for $\mathrm{GL}_n(F)$ is *essentially tame* if its restriction to the wild inertia subgroup $\mathbf{P}_F$ of $\mathbf{W}_F$ is a direct sum of characters. Clearly $\overline{\phi}$ is essentially tame if and only if $\overline{\phi}(\mathbf{P}_F)$ lies in a maximal torus of $\mathrm{GL}_n(\mathbb{C})$, which in turn is equivalent to $\phi(\mathbf{P}_F)$ lying in a maximal torus of $\mathrm{PGL}_n(\mathbb{C})$.

Definition 3.7. A Langlands parameter ϕ for an inner form of $\mathrm{SL}_n(F)$ is essentially tame if $\phi(\mathbf{P}_F)$ lies in a maximal torus of $\mathrm{PGL}_n(\mathbb{C})$.

An important difference with Example 3.5 is that such a projective representation of $\mathbf{P}_F$ can be lifted to a n-dimensional representation of $\mathbf{P}_F$ with the same depth. By [BuHe1, Corollary A.4] any L-parameter for (an inner form of) $\mathrm{GL}_n(F)$ is essentially tame if the residual characteristic of F does not divide n. Our definition is such that the same holds for Langlands parameters for (inner forms of) $\mathrm{SL}_n(F)$.

For such L-parameters the LLC does preserve depths:

Theorem 3.8. *Let $\phi \in \Phi(SL_n(F))$ be essentially tame and relevant for $GL_m(D)_{der}$. Then $d(\pi) = d(\phi)$ for every $\pi \in \Pi_\phi(GL_m(D)_{der})$.*

Proof. Let $\overline{\phi}$ be as in (38), so $d(\overline{\phi}) = d(\pi)$.

First we consider the case where $\overline{\phi}$ is an irreducible n-dimensional representation of $\mathbf{W}_F$. As $\overline{\phi}$ is essentially tame, [BuHe1, Theorem A.3] shows that there exist a finite, tamely ramified Galois extension E/F and a smooth character $\xi : \mathbf{W}_E \to \mathbb{C}^\times$ such that $\overline{\phi} = \mathrm{ind}_{\mathbf{W}_E}^{\mathbf{W}_F} \xi$. We may and will assume that E is contained in our chosen separable closure F_s of F. By Mackey's induction–restriction formula

$$\mathrm{Res}_{\mathbf{W}_E}^{\mathbf{W}_F}(\overline{\phi}) = \bigoplus\nolimits_{s \in \mathbf{W}_F/\mathbf{W}_E} \xi^s \text{ , where } \quad \xi^s(w) = \xi(s^{-1}ws).$$

The elements of $\mathbf{W}_F \setminus \mathbf{W}_E$ permute the $\mathbf{W}_E$-subrepresentations ξ^s nontrivially, so they cannot lie in the kernel of $\overline{\phi}$:

$$\ker \overline{\phi} = \mathbf{W}_E \cap \ker \overline{\phi} = \{w \in \mathbf{W}_E : \xi^s(w) = 1 \ \forall s \in \mathbf{W}_F\}.$$

Let pr: $\mathrm{GL}_n(\mathbb{C}) \to \mathrm{PGL}_n(\mathbb{C})$ be the canonical projection. Then $\phi = \mathrm{pr} \circ \overline{\phi}$ and

$$\ker \phi = \overline{\phi}^{-1}\big(Z(\mathrm{GL}_n(\mathbb{C}))\big) = \{w \in \mathbf{W}_E : \xi^s(w) = \xi(w) \ \forall s \in \mathbf{W}_F\}. \tag{40}$$

Suppose that $d(\overline{\phi}) > d(\phi)$. In view of the definition of $d(\phi)$,

$$\ker \overline{\phi} \supset \mathbf{W}_E \cap \mathrm{Gal}(F_s/F)^{d(\overline{\phi})+}, \text{ but } \ker \overline{\phi} \not\supset \mathbf{W}_E \cap \mathrm{Gal}(F_s/F)^{d(\overline{\phi})} \subset \ker \phi. \tag{41}$$

The relation between the upper and the lower numbering of the filtration subgroups of $\mathbf{W}_F$ [Ser, §IV.3], combined with the compatibility of the lower numbering with subgroups [Ser, Proposition IV.1.2], provides a $l \in \mathbb{R}_{\geq 0}$ such that

$$\mathbf{W}_E \cap \mathrm{Gal}(F_s/F)^{d(\overline{\phi})} = \mathrm{Gal}(F_s/E)^l. \tag{42}$$

In fact $l > 0$ because $d(\overline{\phi}) > d(\phi) \geq 0$. Since $\mathrm{Res}_{\mathbf{W}_E}^{\mathbf{W}_F} \overline{\phi}$ is a direct sum of characters, it factors through the Artin reciprocity map $\mathbf{a}_E : \mathbf{W}_E \to E^\times$. With (41) we see that

$$\mathbf{a}_E(\mathrm{Gal}(F_s/E)^l) \neq \mathbf{a}_E(\mathrm{Gal}(F_s/E)^{l+}).$$

By Theorem 3.6 applied to F_s/E, l must be a positive integer. When we transfer the conjugation action of $\mathbf{W}_F$ on $\mathbf{W}_E$ to $E^\times$ via Artin reciprocity, it becomes the standard action of $\mathrm{Gal}(E/F) \cong \mathbf{W}_F/\mathbf{W}_E$ on $E^\times$. Now (40) says that ξ is a $\mathrm{Gal}(E/F)$-invariant character of U_E^l. Since $l \in \mathbb{Z}_{>0}$ and E/F is tamely ramified, U_E^l is a cohomologically trivial $\mathrm{Gal}(E/F)$-module. According to [BuHe1, Lemma A.1], these properties imply that ξ factors through the norm map $\mathrm{N}_{E/F}$, and there is a unique smooth character

$$\xi' \text{ of } U_E^l \cap F^\times = \mathrm{N}_{E/F}(U_E^l) \text{ such that } \xi = \xi' \circ \mathrm{N}_{E/F} \text{ on } U_E^l.$$

Since $F^\times/\ker(\xi')$ is a finitely generated abelian group and $\mathbb{C}^\times$ is divisible, we can extend ξ' to a smooth character of $F^\times$. Via Artin reciprocity this yields a character ξ_F of $\mathbf{W}_F$. From (42) and the commutative diagram [Ser, §XI.3]

$$\begin{array}{ccc} \mathbf{W}_E & \longrightarrow & \mathbf{W}_F \\ \mathbf{a}_E \big\downarrow & & \big\downarrow \mathbf{a}_F \\ E^\times & \xrightarrow{\mathrm{N}_{E/F}} & F^\times \end{array}$$

we see that $\xi_F = \xi$ on $\mathrm{Gal}(F_s/E)^{d(\overline{\phi})} \cap \mathbf{W}_E$. Then $\overline{\phi} \otimes \xi_F^{-1}$ is another lift of ϕ to $\Phi(\mathrm{GL}_n(F))$, and $\ker \overline{\phi} \otimes \xi_F^{-1}$ contains $\mathrm{Gal}(F_s/E)^{d(\overline{\phi})}$. Thus $d(\overline{\phi} \otimes \xi_F^{-1}) < d(\overline{\phi})$, which contradicts the definition of $\overline{\phi}$. We have shown that $d(\overline{\phi}) = d(\phi)$ if $\overline{\phi}|_{\mathbf{W}_F}$ is irreducible.

For a general essentially tame parameter $\overline{\phi}$ for $\mathrm{GL}_n(F)$, $\overline{\phi}|_{\mathbf{W}_F}$ is a direct sum of irreducible essentially tame parameters ψ_i for $\mathrm{GL}_{n_i}(F)$, with $n_i \leq n$. Writing $\psi_i = \mathrm{ind}_{\mathbf{W}_{E_i}}^{\mathbf{W}_F} \xi_i$ as above, we obtain from (40) that

$$\ker \phi = \{w \in \cap_i \mathbf{W}_{E_i} : \xi_i^s(w) = \xi_j(w) \text{ for all } i,j \text{ and all } s \in \mathbf{W}_F\}. \tag{43}$$

In general this is smaller than

$$\ker(\oplus_i \mathrm{pr}_i \circ \psi_i) = \{w \in \cap_i \mathbf{W}_{E_i} : \xi_i^s(w) = \xi_i(w) \text{ for all } i \text{ and all } s \in \mathbf{W}_F\},$$

where $\mathrm{pr}_i : \mathrm{GL}_{n_i}(\mathbb{C}) \to \mathrm{PGL}_{n_i}(\mathbb{C})$ denotes the canonical projection. Comparing all these kernels we deduce that

$$\max_i d(\psi_i) = d(\overline{\phi}) \geq d(\phi) \geq d(\oplus_i \mathrm{pr}_i \circ \psi_i) = \max_i d(\mathrm{pr}_i \circ \psi_i). \tag{44}$$

However, we cannot just twist $\overline{\phi}$ with a character of $\mathbf{W}_F$ derived from the most ramified of the ψ_i as in the irreducible case, because that could make the depth of another ψ_j much larger.

We suppose once again that $d(\overline{\phi}) > d(\phi)$. Then

$$\ker \overline{\phi} \supset \mathrm{Gal}(F_s/F)^{d(\overline{\phi})+}, \text{ but } \ker \overline{\phi} \not\supset \mathrm{Gal}(F_s/F)^{d(\overline{\phi})} \subset \ker \phi.$$

By (43) all the ξ_i agree on $\mathrm{Gal}(F_s/F)^{d(\overline{\phi})}$. The above method produces characters ξ_i' of $U_{E_i}^{d_i} \cap F^\times$, which agree on $\mathbf{a}_F\big(\mathbf{W}_F^{d(\overline{\phi})}\big)$. Put $\xi' = \xi_i'|_{\mathbf{a}_F\left(\mathbf{W}_F^{d(\overline{\phi})}\right)}$. Now the same argument as in the irreducible case leads to a contradiction with (38). □

References

[ABPS1] A.-M. Aubert, P. Baum, R.J. Plymen, M. Solleveld, Geometric structure and the local Langlands conjecture (2013). arXiv:1211.0180

[ABPS2] A.-M. Aubert, P.F. Baum, R.J. Plymen, M. Solleveld, The local Langlands correspondence for inner forms of SL_n. Res. Math. Sci. (2013, To appear). Preprint

[Bad] A.I. Badulescu, Correspondance de Jacquet-Langlands pour les corps locaux de caractéristique non nulle. Ann. Sci. Éc. Norm. Sup. (4) **35**, 695–747 (2002)

[BaBr] A.I. Badulescu, Un théorème de finitude (with an appendix by P. Broussous). Compos. Math. **132**, 177–190 (2002)

[Bor] A. Borel, Automorphic L-functions. Proc. Symp. Pure Math **33**(2), 27–61 (1979)

[Bro] P. Broussous, Minimal strata for $\mathrm{GL}_m(D)$. J. Reine Angew. Math. **514**, 199–236 (1999)

[BrLe] P. Broussous, B. Lemaire, Building of $\mathrm{GL}(m, D)$ and centralizers. Transform. Groups **7**(1), 15–50 (2002)

[BuFr1] C.J. Bushnell, A. Frölich, *Gauss Sums and p-adic Division Algebras*. Lecture Notes in Mathematics, vol. 987 (Springer, Berlin, 1983)

[BuFr2] C.J. Bushnell, A. Frölich, Non-abelian congruence Gauss sums and p-adic simple algebras. Proc. Lond. Math. Soc. (3) **50**, 207–264 (1985)

[BuHe1] C.J. Bushnell, G. Henniart, The essentially tame local Langlands correspondence, I. J. Am. Math. Soc. **18.3**, 685–710 (2005)

[BuHe2] C.J. Bushnell, G. Henniart, *The Local Langlands Conjecture for GL*(2). Grundlehren der mathematischen Wissenschaften, vol. 335 (Springer, Berlin, 2006)

[BuKu] C.J. Bushnell, P.C. Kutzko, The admissible dual of $\mathrm{SL}(N)$ I. Ann. Scient. Éc. Norm. Sup. (4) **26**, 261–280 (1993)

[ChKa] T.H. Chen, M. Kamgarpour, Preservation of depth in local geometric Langlands correspondence (2014). arXiv: 1404.0598

[DBRe] S. DeBacker, M. Reeder, Depth zero supercuspidal L-packets and their stability. Ann. Math. **169**(3), 795–901 (2009)

[DKV] P. Deligne, D. Kazhdan, M.-F. Vigneras, Représentations des algèbres centrales simples p-adiques, in *Représentations des Groupes réductifs sur un Corps Local* (Travaux en cours, Hermann, 1984), pp. 33–117

[Gan] R. Ganapathy, The local Langlands correspondence for GSp_4 over local function fields. Am. J. Math. **137**(6), 1441–1534 (2015)

[GoJa] R. Godement, H. Jacquet, *Zeta Functions of Simple Algebras*. Lecture Notes in Mathematics, vol. 260 (Springer, New York, 1972)

[GSZ] M. Grabitz, A.J. Silberger, W. Zink, Level zero types and Hecke algebras for local central simple algebras. J. Number Theory **91**, 92–125 (2001)

[GrRe] B.H. Gross, M. Reeder, "Arithmetic invariants of discrete Langlands parameters", Duke Math. J. **154** (2010), 431–508.

[HaTa] M. Harris, R. Taylor, *The Geometry and Cohomology of Some Simple Shimura Varieties*. Annals of Mathematics Studies, vol. 151 (Princeton University Press, Princeton NJ, 2001)

[Hen] G. Henniart, On the local Langlands conjecture for GL(n): the cyclic case. Ann. Math. **123**, 143–203 (1986)

[Hen2] G. Henniart, Une preuve simple de conjectures de Langlands pour GL(n) sur un corps p-adique. Invent. Math. **139** (2000), 439–455.

[HiSa] K. Hiraga, H. Saito, On L-packets for inner forms of SL_n. Mem. Am. Math. Soc. **215**(1013) (2012)

[Jac] H. Jacquet, Principal L-functions of the linear group. Proc. Symp. Pure Math. **33**(2), 63–86 (1979)

[JaLa] H. Jacquet, R. Langlands, *Automorphic Forms on GL*(2). Lecture Notes in Mathematics, vol. 114 (Springer, New York, 1970)

[Kot] R.E. Kottwitz, Stable trace formula: cuspidal tempered terms. Duke Math. J. **51**(3), 611–650 (1984)

[LaRa] J. Lansky, A. Raghuram, On the correspondence of representations between GL(n) and division algebras. Proc. Am. Math. Soc. **131**(5), 1641–1648 (2002)

[LRS] G. Laumon, M. Rapoport, U. Stuhler, $\mathcal{D}$-elliptic sheaves and the Langlands correspondence. Invent. Math. **113**, 217–238 (1993)

[Lus1] G. Lusztig, Classification of unipotent representations of simple p-adic groups. Int. Math. Res. Not. **11**, 517–589 (1995)

[Lus2] G. Lusztig, Classification of unipotent representations of simple p-adic groups II. Represent. Theory **6**, 243–289 (2002)

[Mœ] C. Mœglin, Stabilité en niveau 0, pour les groupes orthogonaux impairs p-adiques. Doc. Math. **9**, 527–564 (2004)

[Mor] L. Morris, Tamely ramified intertwining algebras. Invent. Math. **114**(1), 1–54 (1993)

[MoPr1] A. Moy, G. Prasad, Unrefined minimal K-types for p-adic groups. Invent. Math. **116**, 393–408 (1994)

[MoPr2] A. Moy, G. Prasad, Jacquet functors and unrefined minimal K-types. Comment. Math. Helv. **71**, 98–121 (1996)

[Ree] M. Reeder, Supercuspidal L-packets of positive depth and twisted Coxeter elements. J. Reine Angew. Math. **620**, 1–33 (2008)

[ReYu] M. Reeder, J.-K. Yu, Epipelagic representations and invariant theory. J. Am. Math. Soc. **27**, 437–477 (2014)

[SéSt1] V. Sécherre, S. Stevens, Représentations lisses de $GL_m(D)$ IV: représentations supercuspidales. J. Inst. Math. Jussieu **7**(3), 527–574 (2008)

[Ser] J.-P. Serre, *Corps Locaux* (Hermann, Paris, 1962)

[Tad] M. Tadić, Induced representations of GL(n,A) for p-adic division algebras A. J. Reine Angew. Math. **405**, 48–77 (1990)

[Tat] J. Tate, Number theoretic background. Proc. Symp. Pure Math **33**(2), 3–26 (1979)

[Vog] D. Vogan, The local Langlands conjecture, in *Representation Theory of Groups and Algebras*. Contemporary Mathematics, vol. 145 (American Mathematical Society, Providence, RI, 1993), pp. 305–379

[Wei] A. Weil, Exercices dyadiques. Invent. Math. **27**, 1–22 (1974)

[Yu1] J.-K. Yu, Bruhat-Tits theory and buildings, in *Ottawa Lectures on Admissible Representations of Reductive p-adic Groups*. Fields Institute Monographs (American Mathematical Society, Providence, RI, 2009), pp. 53–77

[Yu2] J.-K. Yu, On the local Langlands correspondence for tori, in *Ottawa Lectures on Admissible Representations of Reductive p-adic Groups*. Fields Institute Monographs (American Mathematical Society, Providence, RI, 2009), pp. 177–183

[Zel] A.V. Zelevinsky, Induced representations of reductive p-adic groups II. On irreducible representations of GL(n). Ann. Sci. École Norm. Sup. (4) **13**(2), 165–210 (1980)

Guide to Elliptic Boundary Value Problems for Dirac-Type Operators

Christian Bär and Werner Ballmann

Dedicated to the memory of Friedrich Hirzebruch

Abstract We present an introduction to boundary value problems for Dirac-type operators on complete Riemannian manifolds with compact boundary. We introduce a very general class of boundary conditions which contain local elliptic boundary conditions in the sense of Lopatinski and Shapiro as well as the Atiyah–Patodi–Singer boundary conditions. We discuss boundary regularity of solutions and also spectral and index theory. The emphasis is on providing the reader with a working knowledge.

Keywords Operators of Dirac type • Boundary conditions • Boundary regularity • Coercivity • Coercivity at infinity • Spectral theory • Index theory • Decomposition theorem • Relative index theorem • Cobordism theorem

2010 *Mathematics Subject Classification.* 35J56, 58J05, 58J20, 58J32.

1 Introduction

Boundary value problems for elliptic differential equations of second order, such as the Dirichlet problem for harmonic functions, have been the object of intense investigation since the nineteenth century. For a large class of such problems, the analysis is by now classical and well understood. There are numerous applications in and outside mathematics.

C. Bär
Institut für Mathematik, Universität Potsdam, Karl-Liebenecht-Str. 24-25, Haus 9, 14476 Potsdam, Germany
e-mail: baer@math.uni-potsdam.de

W. Ballmann (✉)
Max Planck Institute for Mathematics, Vivatsgasse 7, 53111 Bonn, Germany
e-mail: hwbllmnn@mpim-bonn.mpg.de

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_3

The situation is much less satisfactory for boundary value problems for first-order elliptic differential operators such as the Dirac operator. Let us illustrate the phenomena that arise with the elementary example of holomorphic functions on the closed unit disk $D \subset \mathbb{C}$. Holomorphic functions are the solutions of the elliptic equation $\bar{\partial} f = 0$. The real and imaginary parts of f are harmonic and they determine each other up to a constant. Thus for most smooth functions $g : \partial D \to \mathbb{C}$, the Dirichlet problem $\bar{\partial} f = 0$, $f|_{\partial D} = g$, is not solvable. Hence such a boundary condition is too strong for first-order operators.

Ideally, a "good" boundary condition should ensure that the equation $\bar{\partial} f = h$ has a unique solution for given h. At least we want to have that the kernel and the cokernel of $\bar{\partial}$ become finite dimensional, more precisely, that $\bar{\partial}$ becomes a Fredholm operator. If we expand the boundary values of f in a Fourier series, $f(e^{it}) = \sum_{k=-\infty}^{\infty} a_k e^{ikt}$, then we see $a_{-1} = a_{-2} = \cdots = 0$ because otherwise f would have a pole at $z = 0$. Therefore it suffices to impose $a_0 = a_1 = a_2 = \cdots = 0$ to make the kernel trivial. Similarly, imposing $a_k = a_{k+1} = a_{k+2} = \cdots = 0$ would make the kernel k-dimensional. These are typical examples for the nonlocal boundary conditions that one has to consider when dealing with elliptic operators of first order.

A major break-through towards a general theory was achieved in the seminal article [APS], where Atiyah, Patodi, and Singer obtain an index theorem for a certain class of first-order elliptic differential operators on compact manifolds with boundary. This work lies at the heart of many investigations concerning boundary value problems and L^2-index theory for first-order elliptic differential operators.

The aim of the present paper is to provide an introduction to the general theory of elliptic boundary value problems for Dirac-type operators and to give the reader a sound working knowledge of this material. To a large extent, we follow [BB] where all details are worked out but, due to its length and technical complexity, that article may not be a good first start. Results which we only cite here are marked by a ■. The present paper also contains new additions to the results in [BB]; they are given full proofs, terminated by a □. For previous results and alternative approaches see the list of references in [BB].

After some preliminaries on differential operators in Sect. 2, we discuss Dirac-type operators in Sect. 3. An important class consists of Dirac operators in the sense of Gromov and Lawson [GL, LM] associated with Dirac bundles. In Sect. 4, we introduce boundary value problems for Dirac-type operators as defined in [BB]. We discuss their regularity theory. For instance, Theorem 4.9 applied to $\bar{\partial}$ tells us, that, for given $h \in C^\infty(D, \mathbb{C})$, any solution f of $\bar{\partial} f = h$ satisfying the boundary conditions described above will be smooth up to the boundary. We explain that the classical examples, like local elliptic boundary conditions in the sense of Lopatinski and Shapiro and the boundary conditions introduced by Atiyah, Patodi, and Singer, belong to our class of boundary value problems. This class also contains examples which cannot be described by pseudo-differential operators. In Sect. 5, we investigate the spectral theory associated with boundary conditions. The index theory for boundary value problems is the topic of Sect. 6. In general, we assume that the underlying manifold M is a complete, not necessarily compact, Riemannian manifold with compact boundary. We discuss coercivity conditions which ensure the Fredholm property also for noncompact M.

2 Preliminaries

Let M be a Riemannian manifold with compact boundary ∂M and interior unit normal vector field ν along ∂M. The Riemannian volume element on M will be denoted by dV, the one on ∂M by dS. Denote the interior part of M by $\mathring{M}$.

For a vector bundle E over M denote by $C^\infty(M,E)$ the space of smooth sections of E and by $C_c^\infty(M,E)$ and $C_{cc}^\infty(M,E)$ the subspaces of $C^\infty(M,E)$ which consist of smooth sections with compact support in M and $\mathring{M}$, respectively. Let $L^2(M,E)$ be the Hilbert space (of equivalence classes) of square-integrable sections of E and $L^2_{\mathrm{loc}}(M,E)$ be the space of locally square-integrable sections of E. For any integer $k \geq 0$, denote by $H^k_{\mathrm{loc}}(M,E)$ the space of sections of E which have weak derivatives up to order k (with respect to some or any connection on E) that are locally square-integrable.

2.1 *Differential Operators*

Let E and F be Hermitian vector bundles over M and

$$D : C^\infty(M,E) \to C^\infty(M,F)$$

be a differential operator of order (at most) ℓ from E to F. For simplicity, we only consider the case of complex vector bundles. If D acts on real vector bundles one can complexify and thus reduce to the complex case.

Denote by D^* the *formal adjoint* of D. This is the unique differential operator of order (at most) ℓ from F to E such that

$$\int_M \langle D\Phi, \Psi \rangle \,\mathrm{dV} = \int_M \langle \Phi, D^*\Psi \rangle \,\mathrm{dV},$$

for all $\Phi \in C_{cc}^\infty(M,E)$ and $\Psi \in C^\infty(M,F)$. We say that D is *formally self-adjoint* if $E = F$ and $D = D^*$.

Consider D as an unbounded operator, D_{cc}, from $L^2(M,E)$ to $L^2(M,F)$ with domain $\operatorname{dom} D_{cc} = C_{cc}^\infty(M,E)$, and similarly for D^*. The *minimal extension* $D_{\min}$ of D is obtained by taking the closure of the graph of D_{cc} in $L^2(M,E) \oplus L^2(M,F)$. In other words, $\Phi \in L^2(M,E)$ belongs to the domain $\operatorname{dom} D_{\min}$ of $D_{\min}$ if there is a sequence (Φ_n) in $C_{cc}^\infty(M,E)$ which converges to Φ in $L^2(M,E)$ such that $(D\Phi_n)$ is a Cauchy sequence in $L^2(M,F)$; then we set $D_{\min}\Phi := \lim_n D\Phi_n$. By definition, $C^\infty_{cc}(M,E)$ is dense in $\operatorname{dom} D_{\min}$ with respect to the graph norm of $D_{\min}$. The *maximal extension* $D_{\max}$ of D is defined to be the adjoint operator of D^*_{cc}, that is, Φ in $L^2(M,E)$ belongs to the domain $\operatorname{dom} D_{\max}$ of $D_{\max}$ if there is a section $\Xi \in L^2(M,F)$ such that $D\Phi = \Xi$ in the sense of distributions:

$$\int_M \langle \Xi, \Psi\rangle \, \mathrm{dV} = \int_M \langle \Phi, D^*\Psi\rangle \, \mathrm{dV},$$

for all $\Psi \in C^\infty_{cc}(M,F)$; then we set $D_{\max}\Phi := \Xi$. In other words, $(\Phi, -\Xi)$ is perpendicular to the graph of D^*_{cc} in $L^2(M,E) \oplus L^2(M,F)$. Equivalently, $(\Phi, -\Xi)$ is perpendicular to the graph of $D^*_{\min}$ in $L^2(M,E) \oplus L^2(M,F)$. It is easy to see that

$$D_{\min} \subset D_{\max}$$

in the sense that $\operatorname{dom} D_{\min} \subset \operatorname{dom} D_{\max}$ and $D_{\max}|_{\operatorname{dom} D_{\min}} = D_{\min}$. By definition, $D_{\min}$ and $D_{\max}$ are *closed operators*, meaning that their graphs are closed subspaces of $L^2(M,E) \oplus L^2(M,F)$. Hence the *graph norm*, that is, the norm associated with the scalar product

$$(\Phi, \Psi)_D := \int_M (\langle \Phi, \Psi\rangle + \langle D_{\max}\Phi, D_{\max}\Psi\rangle)\mathrm{dV},$$

turns $\operatorname{dom} D_{\min}$ and $\operatorname{dom} D_{\max}$ into Hilbert spaces. Boundary value problems in our sense are concerned with closed operators lying between $D_{\min}$ and $D_{\max}$.

2.2 The Principal Symbol

For a differential operator D from E to F of order (at most) ℓ as above, there is a field $\sigma_D : (T^*M)^\ell \to \operatorname{Hom}(E,F)$ of symmetric ℓ-linear maps, the *principal symbol* σ_D of D, defined by the ℓ-fold commutator[1]

$$\sigma_D(df_1, \dots, df_\ell) := \frac{1}{\ell!}[\dots[D,f_1], \dots, f_\ell],$$

for all $f_1, \dots, f_\ell \in C^\infty(M,\mathbb{R})$. In the case $\ell = 1$, this means that

$$D(f\Phi) = \sigma_D(df)\Phi + fD\Phi,$$

for all $f \in C^\infty(M,\mathbb{R})$ and $\Phi \in C^\infty(M,E)$. The principal symbol σ_D vanishes precisely at those points where the order of D is at most $\ell - 1$. The principal symbol of D^* is

[1] Here $[D,f] = D \circ (f \cdot \mathrm{id}_E) - (f \cdot \mathrm{id}_F) \circ D$.

$$\sigma_{D^*}(\xi_1, \dots, \xi_k) = (-1)^\ell \sigma_D(\xi_1, \dots, \xi_\ell)^*, \tag{1}$$

for all $\xi_1, \dots, \xi_\ell \in T^*M$. Since σ_D is symmetric in $\xi_1, \dots, \xi_\ell$, it is determined by its values along the diagonal; we use $\sigma_D(\xi)$ as a shorthand notation for $\sigma_D(\xi, \dots, \xi)$. Then we have, for all $\xi \in T^*M$,

$$\sigma_{D_1 D_2}(\xi) = \sigma_{D_1}(\xi) \circ \sigma_{D_2}(\xi) \tag{2}$$

for the principal symbol of the composition of differential operators D_1 of order ℓ_1 and D_2 of order ℓ_2.

The Riemannian metric induces a vector bundle isomorphism $TM \to T^*M$, $X \mapsto X^\flat$, defined by $\langle X, Y \rangle = X^\flat(Y)$ for all Y. The inverse isomorphism $T^*M \to TM$ is denoted by $\xi \mapsto \xi^\sharp$.

Proposition 2.1 (Green's Formula). *Let D be a differential operator from E to F of order one. Then we have, for all $\Phi \in C_c^\infty(M, E)$ and $\Psi \in C_c^\infty(M, F)$,*

$$\int_M \langle D\Phi, \Psi \rangle \, dV = \int_M \langle \Phi, D^*\Psi \rangle \, dV - \int_{\partial M} \langle \sigma_D(\nu^\flat)\Phi, \Psi \rangle \, dS \,. \qquad \blacksquare$$

For a proof see, e.g., [Ta, Proposition 9.1, p. 160].

Examples 2.2. By definition, a connection ∇ on E is a differential operator from E to $T^*M \otimes E$ of order one such that $[\nabla, f](\Phi) = df \otimes \Phi$. We obtain

$$\sigma_\nabla(\xi)(\Phi) = \xi \otimes \Phi \quad \text{and} \quad \sigma_{\nabla^*}(\xi)(\Psi) = -\Psi(\xi^\sharp). \tag{3}$$

Hence all connections on E have the same principal symbol reflecting the fact that the difference of two connections is of order zero.

There are two natural differential operators of order two associated with ∇, the second covariant derivative ∇^2 with principal symbol

$$\sigma_{\nabla^2}(\xi)(\Phi) = \xi \otimes \xi \otimes \Phi \tag{4}$$

and the connection Laplacian $\nabla^*\nabla$ with principal symbol

$$\sigma_{\nabla^*\nabla}(\xi)(\Phi) = -|\xi|^2\Phi, \tag{5}$$

and both, (4) and (5), are in agreement with (2) and (3).

2.3 *Elliptic Operators*

We say that D is *elliptic* if $\sigma_D(\xi) : E_x \to F_x$ is an isomorphism, for all $x \in M$ and nonzero $\xi \in T_x^*M$. In the above examples, ∇, ∇^*, and ∇^2 are not elliptic; in fact, the involved bundles have different rank. On the other hand, the connection Laplacian is elliptic, by (5).

Suppose that D is elliptic. Then *interior elliptic regularity* says that, for any given integer $k \geq 0$, $\Phi \in \operatorname{dom} D_{\max}$ is contained in $H^{k+\ell}_{\mathrm{loc}}(\mathring{M}, E)$ if $D_{\max}\Phi$ belongs to $H^k_{\mathrm{loc}}(\mathring{M}, F)$. In particular, if $\Phi \in \operatorname{dom} D_{\max}$ satisfies $D_{\max}\Phi \in C^\infty(\mathring{M}, F)$, then $\Phi \in C^\infty(\mathring{M}, E)$.

If M is closed and D is elliptic and formally self-adjoint, then the eigenspaces of D are finite dimensional, contained in $C^\infty(M, E)$, pairwise perpendicular with respect to the L^2-product, and span $L^2(M, E)$. As an example, the connection Laplacian is elliptic and formally self-adjoint.

For any differential operator $D : C^\infty(M, E) \to C^\infty(M, F)$ of order one, consider the fiberwise linear bundle map

$$\mathscr{A}_D : T^*M \otimes \operatorname{Hom}(E, E) \to \operatorname{Hom}(E, F), \quad V \mapsto \sum_j \sigma_D(e_j^*) \circ V(e_j).$$

Here $(e_1, \ldots, e_n)$ is any local tangent frame and $(e_1^*, \ldots, e_n^*)$ its associated dual cotangent frame of M. Note that $\mathscr{A}_D$ does not depend on the choice of frame.

Proposition 2.3. *Let $D : C^\infty(M, E) \to C^\infty(M, F)$ be a differential operator of order one such that $\mathscr{A}_D$ is onto. Then there exists a connection ∇ on E such that*

$$D = \sum_j \sigma_D(e_j^*) \circ \nabla_{e_j},$$

for any local tangent frame $(e_1, \ldots, e_n)$ and the associated dual cotangent frame $(e_1^*, \ldots, e_n^*)$ of M.

The proof can be found in Appendix 2.

If D is elliptic, $\mathscr{A}_D$ is onto: given $U \in \operatorname{Hom}(E, F)$ put $V(e_2) = \ldots = V(e_n) = 0$ and $V(e_1) = \sigma_D(e_1^*)^{-1} \circ U$, for instance. Hence Proposition 2.3 applies and we have

Corollary 2.4. *Let $D : C^\infty(M, E) \to C^\infty(M, F)$ be an elliptic differential operator of order one. Then there exists a connection ∇ on E such that*

$$D = \sum_j \sigma_D(e_j^*) \circ \nabla_{e_j},$$

for any local tangent frame $(e_1, \ldots, e_n)$ and the associated dual cotangent frame $(e_1^*, \ldots, e_n^*)$ of M. □

In the special case of Dirac-type operators (see definition below), this corollary is [AT, Lemma 2.1]. Proposition 2.3 is also useful for nonelliptic operators. For instance, it applies to Dirac-type operators on Lorentzian manifolds; these are hyperbolic instead of elliptic.

3 Dirac-Type Operators

From now on we concentrate on an important special class of first-order elliptic operators.

3.1 Clifford Relations and Dirac-Type Operators

We say that a differential operator $D : C^\infty(M, E) \to C^\infty(M, F)$ of order one is of *Dirac type* if its principal symbol σ_D satisfies the *Clifford relations*,

$$\sigma_D(\xi)^* \sigma_D(\eta) + \sigma_D(\eta)^* \sigma_D(\xi) = 2 \langle \xi, \eta \rangle \cdot \mathrm{id}_{E_x}, \tag{6}$$

$$\sigma_D(\xi) \sigma_D(\eta)^* + \sigma_D(\eta) \sigma_D(\xi)^* = 2 \langle \xi, \eta \rangle \cdot \mathrm{id}_{F_x}, \tag{7}$$

for all $x \in M$ and $\xi, \eta \in T_x^* M$.

The classical Dirac operator on a spin manifold is an important example. More generally, the class of Dirac-type operators contains Dirac operators on Dirac bundles as in [LM, Chap. II, § 5].

By (1), if D is of Dirac type, then so is D^*. Furthermore, by (6) and (7), Dirac-type operators are elliptic with

$$\sigma_D(\xi)^{-1} = |\xi|^{-2} \sigma_D(\xi)^*, \text{ for all nonzero } \xi \in T^* M. \tag{8}$$

If D is a formally self-adjoint operator of Dirac type on E, then the endomorphisms $\sigma_D(\xi)$ are skewhermitian, $\xi \in T^* M$. In this case, the Clifford relations (6) and (7) may be spelled out as

$$\sigma_D(\xi) \sigma_D(\eta) + \sigma_D(\eta) \sigma_D(\xi) = -2 \langle \xi, \eta \rangle \cdot \mathrm{id}_{E_x},$$

for all $x \in M$ and $\xi, \eta \in T_x^* M$. In other words, the principal symbol turns E into a bundle of modules over the Clifford algebras $\mathrm{Cliff}(T^* M)$.

Proposition 3.1 (Weitzenböck Formula). *Let $D : C^\infty(M, E) \to C^\infty(M, F)$ be of Dirac type. Then there exists a unique metric connection ∇ on E with*

$$D^* D = \nabla^* \nabla + \mathscr{K}, \tag{9}$$

where $\mathscr{K}$ is a field of symmetric endomorphisms of E.

See Appendix 2 for the proof. For special choices for D, this formula is also known as Bochner formula, Bochner–Kodaira formula, or Lichnerowicz formula.

In general, the connections in Corollary 2.4 and Proposition 3.1 do not coincide.

3.2 Adapted Operators on the Boundary

Suppose from now on that D is of Dirac type. For $x \in \partial M$, identify $T_x^*\partial M$ with the space of covectors ξ in T_x^*M such that $\xi(\nu(x)) = 0$. Then, by (6) and (8),

$$\sigma_D(\nu(x)^\flat)^{-1} \circ \sigma_D(\xi) : E_x \to E_x \tag{10}$$

is skewhermitian, for all $x \in \partial M$ and $\xi \in T_x^*\partial M$. Hence there exist formally self-adjoint differential operators $A : C^\infty(\partial M, E) \to C^\infty(\partial M, E)$ of first order with principal symbol

$$\sigma_A(\xi) = \sigma_D(\nu(x)^\flat)^{-1} \circ \sigma_D(\xi). \tag{11}$$

We call such operators *adapted* to D. Note that such an operator A is also of Dirac type and that the zero-order term of A is only unique up to addition of a field of hermitian endomorphisms of E. By (1) and (10) applied to D^*, the principal symbol of an operator $\tilde{A}$ adapted to D^* is

$$\sigma_{\tilde{A}}(\xi) = (-\sigma_D(\nu(x)^\flat)^{-1})^* \circ (-\sigma_D(\xi))^* = \sigma_D(\nu(x)^\flat) \circ \sigma_D(\xi)^*.$$

By (11), this implies

$$\begin{aligned}
\sigma_{\tilde{A}}(\xi) &= \sigma_D(\nu(x)^\flat) \circ (\sigma_D(\nu(x)^\flat) \circ \sigma_A(\xi))^* \\
&= \sigma_D(\nu(x)^\flat) \circ \sigma_A(\xi)^* \circ \sigma_D(\nu(x)^\flat)^* \\
&= \sigma_D(\nu(x)^\flat) \circ \sigma_{-A}(\xi) \circ \sigma_D(\nu(x)^\flat)^{-1}.
\end{aligned}$$

Hence, if A is adapted to D, then

$$\tilde{A} = \sigma_D(\nu^\flat) \circ (-A) \circ \sigma_D(\nu^\flat)^{-1} \tag{12}$$

is adapted to D^*. Given A, this choice of $\tilde{A}$ is the most natural one.

3.3 Formally Self-adjoint Dirac-Type Operators

If the Dirac-type operator D is formally self-adjoint, then there is a particularly useful choice of adapted boundary operator A.

Lemma 3.2. *Let $D : C^\infty(M, E) \to C^\infty(M, E)$ be a formally self-adjoint operator of Dirac type. Then there is an operator A adapted to D along ∂M such that $\sigma_D(\nu^\flat)$ anticommutes with A,*

$$\sigma_D(\nu^\flat) \circ A = -A \circ \sigma_D(\nu^\flat). \tag{13}$$

See Appendix 2 for the proof.

Remarks 3.3. (1) The operator A in Lemma 3.2 is unique up to addition of a field of symmetric endomorphisms of E along ∂M which anticommutes with $\sigma_D(\nu^\flat)$.
(2) If A anticommutes with $\sigma_D(\nu^\flat)$, then $\sigma_D(\nu^\flat)$ induces isomorphisms between the $\pm\lambda$-eigenspaces of A, for all $\lambda \in \mathbb{R}$. In particular, $\ker A$ is invariant under $\sigma_D(\nu^\flat)$ and the η-invariant[2] of A vanishes. Moreover,

$$\omega(\varphi, \psi) := (\sigma_D(\nu^\flat)\varphi, \psi)_{L^2(\partial M)}$$

is a nondegenerate skewhermitian form on $\ker A$ (and also on $L^2(\partial M, E)$).

4 Boundary Value Problems

In this section we will study boundary value problems. This will be done under the following:

Standard Setup 4.1.

- M is a complete Riemannian manifold with compact boundary ∂M;
- ν is the interior unit normal vector field along ∂M;
- E and F are Hermitian vector bundles over M;
- $D : C^\infty(M, E) \to C^\infty(M, F)$ is a Dirac-type operator;
- $A : C^\infty(\partial M, E) \to C^\infty(\partial M, E)$ is a boundary operator adapted to D.

4.1 Spectral Subspaces

If A is adapted to D, then A is a formally self-adjoint elliptic operator over the compact manifold ∂M. Hence we have, in the sense of Hilbert spaces,

$$L^2(\partial M, E) = \oplus_j \mathbb{C} \cdot \varphi_j,$$

where (φ_j) is an orthonormal basis of $L^2(\partial M, E)$ consisting of eigensections of A, $A\varphi_j = \lambda_j \varphi_j$. In terms of such an orthonormal basis, the Sobolev space $H^s(\partial M, E)$, $s \in \mathbb{R}$, consists of all sections

$$\varphi = \sum_j a_j \varphi_j \quad \text{such that} \quad \sum_j |a_j|^2 (1 + \lambda_j^2)^s < \infty,$$

[2]The η-invariant of A is defined as the value of the meromorphic extension of $\eta(s) = \sum_{\lambda \neq 0} \operatorname{sign}(\lambda)|\lambda|^{-s}$ at $s = 0$, see [APS]. Here the sum is taken over all nonzero eigenvalues of A taking multiplicities into account. Hence the η-invariant is a measure for the asymmetry of the spectrum.

where $L^2(\partial M, E) = H^0(\partial M, E)$. The natural pairing

$$H^s(\partial M, E) \times H^{-s}(\partial M, E) \to \mathbb{C}, \quad \big(\sum_j a_j\phi_j, \sum_j b_j\phi_j\big) = \sum_j \bar{a}_j b_j, \tag{14}$$

is perfect, for all $s \in \mathbb{R}$. By the Sobolev embedding theorem,

$$C^\infty(\partial M, E) = \bigcap_{s\in\mathbb{R}} H^s(\partial M, E).$$

Rellich's embedding theorem says that for $s_1 > s_2$ the embedding

$$H^{s_1}(\partial M, E) \hookrightarrow H^{s_2}(\partial M, E)$$

is compact. We also set

$$H^{-\infty}(\partial M, E) := \bigcup_{s\in\mathbb{R}} H^s(\partial M, E).$$

For $I \subset \mathbb{R}$, let Q_I be the associated spectral projection,

$$Q_I : \sum_j a_j\varphi_j \mapsto \sum_{\lambda_j\in I} a_j\varphi_j. \tag{15}$$

Then Q_I is orthogonal and maps $H^s(\partial M, E)$ to itself, for all $s \in \mathbb{R}$. Set

$$H^s_I(A) := Q_I(H^s(\partial M, E)) \subset H^s(\partial M, E).$$

For $a \in \mathbb{R}$, define the hybrid Sobolev spaces

$$\check{H}(A) := H^{1/2}_{(-\infty,a)}(A) \oplus H^{-1/2}_{[a,\infty)}(A), \tag{16}$$

$$\hat{H}(A) := H^{-1/2}_{(-\infty,a)}(A) \oplus H^{1/2}_{[a,\infty)}(A). \tag{17}$$

Note that, as topological vector spaces, $\check{H}(A)$ and $\hat{H}(A)$ do not depend on the choice of a. In particular,

$$\hat{H}(A) = \check{H}(-A).$$

Moreover, the natural pairing

$$\check{H}(A) \times \check{H}(-A) \to \mathbb{C}, \quad \big(\sum_j a_j\phi_j, \sum_j b_j\phi_j\big) = \sum_j \bar{a}_j b_j,$$

is perfect, compare (14).

4.2 The Maximal Domain

Following [BB, Corollary 6.6, Theorem 6.7, Proposition 7.2], we now discuss properties of the maximal domain of D.

Theorem 4.2. *Assume the Standard Setup 4.1. Then the domain of $D_{\max}$, equipped with the graph norm topology, has the following properties:*

(1) $C_c^\infty(M,E)$ is dense in $\operatorname{dom} D_{\max}$;

(2) the trace map $\mathscr{R}\Phi := \Phi|_{\partial M}$ on $C_c^\infty(M,E)$ extends uniquely to a continuous surjection $\mathscr{R} : \operatorname{dom} D_{\max} \to \check{H}(A)$;

(3) $\operatorname{dom} D_{\min} = \{\Phi \in \operatorname{dom} D_{\max} \mid \mathscr{R}\Phi = 0\}$. In particular, $\mathscr{R}$ induces an isomorphism

$$\check{H}(A) \cong \operatorname{dom} D_{\max} / \operatorname{dom} D_{\min};$$

(4) for any closed subspace $B \subset \check{H}(A)$, the operator $D_{B,\max}$ with domain

$$\operatorname{dom} D_{B,\max} = \{\Phi \in \operatorname{dom} D_{\max} \mid \mathscr{R}\Phi \in B\}$$

is a closed extension of D between $D_{\min}$ and $D_{\max}$, and any closed extension of D between $D_{\min}$ and $D_{\max}$ is of this form;

(5) for all $\Phi \in \operatorname{dom} D_{\max}$ and $\Psi \in \operatorname{dom} D^*_{\max}$,

$$\int_M \langle D_{\max}\Phi, \Psi\rangle \mathrm{dV} = \int_M \langle \Phi, D^*_{\max}\Psi\rangle \mathrm{dV} - \int_{\partial M} \langle \sigma_D(\nu^\flat)\mathscr{R}\Phi, \mathscr{R}\Psi\rangle \, \mathrm{dS}\,. \qquad \blacksquare$$

Remark 4.3. As a topological vector space, $\check{H}(A)$ does not depend on the choice of adapted operator A, by Theorem 4.2.3. The pairing in Theorem 4.2.5 is well defined because $\sigma_D(\nu^\flat)$ maps $\check{H}(A)$ to $\hat{H}(A)$ by (12).

Theorem 4.4 (Boundary Regularity I, [BB, Theorem 6.11]). *Assume the Standard Setup 4.1. Let $k \geq 0$ be an integer and $\Phi \in \operatorname{dom} D_{\max}$. Then*

$$\Phi \in H^{k+1}_{\mathrm{loc}}(M,E) \iff D\Phi \in H^k_{\mathrm{loc}}(M,F) \text{ and } Q_{[0,\infty)}\mathscr{R}\Phi \in H^{k+1/2}(\partial M, E).$$

In particular,

$$\Phi \in H^1_{\mathrm{loc}}(M,E) \iff Q_{[0,\infty)}\mathscr{R}\Phi \in H^{1/2}(\partial M, E). \qquad \blacksquare$$

Note that $Q_{[0,\infty)}\mathscr{R}\Phi \in H^{1/2}(\partial M,E)$ if and only if $\mathscr{R}\Phi \in H^{1/2}(\partial M,E)$, by (16) and Theorem 4.2.2.

4.3 Boundary Conditions

Theorem 4.2.4 justifies the following:

Definition 4.5. A *boundary condition* for D is a closed subspace of $\check{H}(A)$.

In the notation of Theorem 4.2.3, we write $D_{B,\max}$ for the operator with boundary values in a boundary condition B. This differs from the notation of Atiyah–Patodi–Singer and others, who would use a projection P with $\ker P = B$ to write $P\mathscr{R}\Phi = 0$.

Theorem 4.6 (The Adjoint Operator, [BB, Sect. 7.2]). *Assume the Standard Setup 4.1 and that $B \subset \check{H}(A)$ is a boundary condition. Let $\tilde{A}$ be adapted to D^*. Then*

$$B^{\mathrm{ad}} := \{\psi \in \check{H}(\tilde{A}) \mid (\sigma_D(\nu^\flat)\phi, \psi) = 0, \text{for all } \phi \in B\}$$

is a closed subspace of $\check{H}(\tilde{A})$, that is, it is a boundary condition for D^*. Moreover, the adjoint operator of $D_{B,\max}$ is the operator $D^*_{B^{\mathrm{ad}},\max}$. ■

4.4 D-Elliptic Boundary Conditions

For $V \subset H^{-\infty}(\partial M, E)$ and $s \in \mathbb{R}$, let

$$V^s := V \cap H^s(\partial M, E)).$$

For subspaces $V, W \subset L^2(\partial M, E)$, we say that a bounded linear operator $g : V \to W$ is of *order zero* if

$$g(V^s) \subset W^s,$$

for all $s \geq 0$. For example, spectral projections Q_I as in (15) are of order zero.

Definition 4.7. A linear subspace $B \subset H^{1/2}(\partial M, E)$ is said to be a *D-elliptic boundary condition* if there is an L^2-orthogonal decomposition

$$L^2(\partial M, E) = V_- \oplus W_- \oplus V_+ \oplus W_+ \tag{18}$$

such that

$$B = W_+ \oplus \{v + gv \mid v \in V_-^{1/2}\},$$

where

(1) W_- and W_+ are finite dimensional and contained in $C^\infty(\partial M, E)$;
(2) $V_- \oplus W_- \subset L^2_{(-\infty,a]}(A)$ and $V_+ \oplus W_+ \subset L^2_{[-a,\infty)}(A)$, for some $a \in \mathbb{R}$;
(3) $g : V_- \to V_+$ and its adjoint $g^* : V_+ \to V_-$ are operators of order 0.

Remarks 4.8. (1) D-elliptic boundary conditions are closed in $\check{H}(A)$, and hence they are boundary conditions in the sense formulated further up.
(2) If B is a D-elliptic boundary condition and $a \in \mathbb{R}$ is given, then the decomposition (18) can be chosen such that

$$V_- \oplus W_- = L^2_{(-\infty,a)}(\partial M, E) \quad \text{and} \quad V_+ \oplus W_+ = L^2_{[a,\infty)}(\partial M, E).$$

(3) If B is a D-elliptic boundary condition, then B^{ad} is D^*-elliptic. In fact, using $\tilde{A}$ as in (12), we get

$$B^{\mathrm{ad}} = \sigma_D(\nu^\flat)\big(W_- \oplus \{v - g^* v \mid v \in V_+^{1/2}\}\big). \tag{19}$$

The point of D-ellipticity of a boundary condition is that it ensures good regularity properties up to boundary just like ellipticity of the differential operator implies good regularity properties in the interior.

Theorem 4.9 (Boundary Regularity II, [BB, Theorem 7.17]). *Assume the Standard Setup 4.1 and that $B \subset \check{H}(A)$ is a D-elliptic boundary condition. Then*

$$\Phi \in H^{k+1}_{\mathrm{loc}}(M, E) \iff D_{B,\max}\Phi \in H^k_{\mathrm{loc}}(M, F),$$

for all $\Phi \in \operatorname{dom} D_{B,\max}$ and integers $k \geq 0$. In particular, $\Phi \in \operatorname{dom} D_{B,\max}$ is smooth up to the boundary if and only if $D\Phi$ is smooth up to the boundary. ■

Theorem 4.10. *Assume the Standard Setup 4.1 and that $B \subset \check{H}(A)$ is a D-elliptic boundary condition. Then*

$$C^\infty_c(M, E; B) := \{\Phi \in C^\infty_c(M, E) \mid \mathscr{R}(\Phi) \in B\}$$

is dense in $\operatorname{dom} D_{B,\max}$ with respect to the graph norm.

Proof. Choose a representation of B as in Remark 4.8.2. Since W_- is finite dimensional and contained in $C^\infty(\partial M, E)$, we get that $V_- \cap C^\infty(\partial M, E)$ is dense in V_-, and similarly for V_+. Since g is of order 0, we conclude that

$$\{v + gv \mid v \in V_-^{1/2}\} \cap C^\infty(\partial M, E)$$

is dense in $\{v + gv \mid v \in V_-^{1/2}\}$. Hence $B \cap C^\infty(\partial M, E)$ is dense in B.

Let $\Phi \in \operatorname{dom} D_{B,\max}$ and set $\varphi := \mathscr{R}\Phi$. Choose an extension operator $\mathscr{E}$ as in (43) in [BB]. Then $\Psi := \Phi - \mathscr{E}\varphi$ vanishes along ∂M, and hence $\Psi \in \operatorname{dom} D_{\min}$, by Theorem 4.2.3. Therefore Ψ is the limit of smooth sections in $C^\infty_{cc}(M, E)$, by the definition of $D_{\min}$.

It remains to show that $\mathscr{E}\varphi$ can be approximated by smooth sections in $C^\infty(M,E;B)$. As explained in the beginning of the proof, there is a sequence (φ_n) in $B \cap C^\infty(\partial M,E)$ converging to φ. Then $\mathscr{E}\varphi_n \in C^\infty(M,E;B)$ and $\mathscr{E}\varphi_n \to \mathscr{E}\varphi$ with respect to the graph norm, by Lemma 5.5 in [BB]. □

4.5 Self-adjoint D-Elliptic Boundary Conditions

Assume the Standard Setup 4.1, that $E = F$ and that D is formally self-adjoint. Choose $\tilde{A}$ as in (12). Let $B \subset H^{1/2}(\partial M,E)$ be a D-elliptic boundary condition. Then $D_{B^{\mathrm{ad}},\max}$ is the adjoint operator of $D_{B,\max}$, where B^{ad} is given by (19). In particular, $D_{B,\max}$ is self-adjoint if and only if B is self-adjoint, that is, if and only if $B = B^{\mathrm{ad}}$.

Note that B^{ad} is the image of the L^2-orthogonal complement of B in $H^{1/2}(\partial M,E)$ under $\sigma_D(\nu^\flat)$. Hence $B = B^{\mathrm{ad}}$ if and only if $\sigma_D(\nu^\flat)$ interchanges B with its L^2-orthogonal complement in $H^{1/2}(\partial M,E)$.

Theorem 4.11. *Assume the Standard Setup 4.1, that $E = F$ and that D is formally self-adjoint. Let B be a self-adjoint D-elliptic boundary condition.*

Then D is essentially self-adjoint on

$$C_c^\infty(M,E;B) = \{\Phi \in C_c^\infty(M,E) \mid \mathscr{R}\Phi \in B\},$$

and the closure of D on $C_c^\infty(M,E;B)$ is $D_{B,\max}$.

Proof. By Theorem 4.10, $C_c^\infty(M,E;B)$ is dense in $\operatorname{dom} D_{B,\max}$. □

The following result adapts and extends Theorem 1.83 in [BBC] to D-elliptic boundary conditions as considered here.

Theorem 4.12 (Normal Form for B). *Assume the Standard Setup 4.1, that $E = F$ and that D is formally self-adjoint. Suppose that $\sigma_D(\nu^\flat)$ anticommutes with A. Then a D-elliptic boundary condition B is self-adjoint if and only if there is*

(1) an orthogonal decomposition $L^2_{(-\infty,0)}(A) = V \oplus W$, where W is a finite dimensional subspace of $C^\infty(\partial M,E)$,

(2) an orthogonal decomposition $\ker A = L \oplus \sigma_D(\nu^\flat)L$,

(3) and a self-adjoint operator $g : V \oplus L \to V \oplus L$ of order zero such that

$$B = \sigma_D(\nu^\flat)W \oplus \{v + \sigma_D(\nu^\flat)gv \mid v \in V^{1/2} \oplus L\}.$$ ■

Remarks 4.13. (1) In Theorem 4.12, the case $\ker A = \{0\}$ is not excluded. In this latter case, the representation of B as in Theorem 4.12 is unique since $V = Q_{(-\infty,0)}B$ and W is the orthogonal complement of V in $L^2_{(-\infty,0)}(A)$.

(2) Theorem 4.12.2 excludes the existence of self-adjoint boundary conditions in the case where $\ker A$ is of odd dimension. Conversely, if $\dim\ker A$ is even and the eigenvalues i and $-i$ of $\sigma_D(\nu^\flat)$ have equal multiplicity, then self-adjoint boundary conditions exist. A simple example is $H^{1/2}_{(-\infty,0)}(A) \oplus L$, where L is a subspace of $\ker A$ as in Theorem 4.12.2.

(3) Let E, D, and A be the complexification of a Riemannian vector bundle, a formally self-adjoint real Dirac-type operator, and a real boundary operator $A_{\mathbb{R}}$, respectively. Then $\sigma_D(\nu^\flat)$ turns the real kernel $\ker(A_{\mathbb{R}})$ into a symplectic vector space. It follows that the complexification L of any Lagrangian subspace of $\ker(A_{\mathbb{R}})$ will satisfy $\ker A = L \oplus \sigma_D(\nu^\flat)L$, and hence self-adjoint elliptic boundary conditions exist, by the previous remark.

(4) First attempts have been made to relax the condition of compactness of ∂M. The results in [GN] apply to the Dirac operator associated with a spin^c structure when M and ∂M are complete and geometrically bounded in a suitable sense.

4.6 Local and Pseudo-Local Boundary Conditions

Throughout this section, we let M be a complete Riemannian manifold with compact boundary, E and F be Hermitian vector bundles over M, and D be a Dirac-type operator from E to F.

Definition 4.14. We say that a linear subspace $B \subset H^{1/2}(\partial M, E)$ is a *local boundary condition* if there is a (smooth) subbundle $E' \subset E_{|\partial M}$ such that

$$B = H^{1/2}(\partial M, E').$$

More generally, we say that B is *pseudo-local* if there is a classical pseudo-differential operator P of order 0 acting on sections of E over ∂M which induces an orthogonal projection on $L^2(\partial M, E)$ such that

$$B = P(H^{1/2}(\partial M, E)).$$

Theorem 4.15 (Characterization of Pseudo-Local Boundary Conditions, [BB, Theorem 7.20]). *Assume the Standard Setup 4.1. Let P be a classical pseudo-differential operator of order zero, acting on sections of E over ∂M. Suppose that P induces an orthogonal projection in $L^2(\partial M, E)$. Then the following are equivalent:*

(i) $B = P(H^{1/2}(\partial M, E))$ is a D-elliptic boundary condition.

(ii) For some (and then all) $a \in \mathbb{R}$,

$$P - Q_{[a,\infty)} : L^2(\partial M, E) \to L^2(\partial M, E)$$

is a Fredholm operator.

(iii) For some (and then all) $a \in \mathbb{R}$,

$$P - Q_{[a,\infty)} : L^2(\partial M, E) \to L^2(\partial M, E)$$

is an elliptic classical pseudo-differential operator of order zero.

(iv) For all $\xi \in T_x^*\partial M \setminus \{0\}$, $x \in \partial M$, the principal symbol $\sigma_P(\xi) : E_x \to E_x$ restricts to an isomorphism from the sum of the eigenspaces for the negative eigenvalues of $i\sigma_A(\xi)$ onto its image $\sigma_P(\xi)(E_x)$. ■

Remark 4.16. The projection P is closely related to the Calderón projector $\mathscr{P}$ studied in the literature, see, e.g., [BW]. If the Calderón projector is chosen selfadjoint as described in [BW, Lemma 12.8], then $P = \mathrm{id} - \mathscr{P}$ satisfies the conditions in Theorem 4.15.

Our concept of D-elliptic boundary conditions covers in particular that of classical elliptic boundary conditions in the sense of Lopatinski and Shapiro.

Corollary 4.17 ([BB, Corollary 7.22]). *Let $E' \subset E_{|\partial M}$ be a subbundle and $P : E_{|\partial M} \to E'$ be the fiberwise orthogonal projection. If $(D, \mathrm{id} - P)$ is an elliptic boundary value problem in the classical sense of Lopatinski and Shapiro, then $B = H^{1/2}(\partial M, E')$ is a local D-elliptic boundary condition.*

Proof. Let $(D, \mathrm{id} - P)$ be an elliptic boundary value problem in the classical sense of Lopatinsky and Shapiro, see, e.g., [Gi, Sect. 1.9]. This means that the rank of E' is half of that of E and that, for any $x \in \partial M$, any $\eta \in T_x^*\partial M \setminus \{0\}$, and any $\phi \in (E_x')^\perp$, there is a unique solution $f : [0, \infty) \to E_x$ to the ordinary differential equation

$$\left(i\sigma_A(\eta) + \frac{d}{dt}\right) f(t) = 0 \qquad (20)$$

subject to the boundary conditions

$$(\mathrm{id} - P)f(0) = \phi \quad \text{and} \quad \lim_{t\to\infty} f(t) = 0.$$

Recall from Sect. 3.2 that $i\sigma_A(\eta)$ is Hermitian, hence diagonalizable with real eigenvalues. The solution to (20) is given by $f(t) = \exp(-it\sigma_A(\eta))\phi$. The condition $\lim_{t\to\infty} f(t) = 0$ is therefore equivalent to ϕ lying in the sum of the eigenspaces to the positive eigenvalues of $i\sigma_A(\eta)$. This shows criterion (iv) of Theorem 4.15. □

As a direct consequence of Theorem 4.15 (iv) we obtain

Corollary 4.18. *Let $E_{|\partial M} = E' \oplus E''$ be a decomposition such that $\sigma_A(\xi) = \sigma_D(\nu^\flat)^{-1}\sigma_D(\xi)$ interchanges E' and E'', for all $\xi \in T^*\partial M$. Then $B' := H^{1/2}(\partial M, E')$ and $B'' := H^{1/2}(\partial M, E'')$ are local D-elliptic boundary conditions.* □

This corollary applies, in particular, if A itself interchanges sections of E' and E''.

4.7 Examples

In this section, we discuss some important elliptic boundary conditions.

Example 4.19 (Differential Forms). Let

$$E = \bigoplus_{j=0}^{n} \Lambda^j T^* M = \Lambda^* T^* M$$

be the sum of the bundles of $\mathbb{C}$-valued alternating forms over M. The Dirac-type operator is given by $D = d + d^*$, where d denotes exterior differentiation.

As before, ν is the interior unit normal vector field along the boundary ∂M and $\nu^\flat$ the associated unit conormal one-form. For each $x \in \partial M$ and $0 \le j \le n$, we have a canonical identification

$$\Lambda^j T_x^* M = \big(\Lambda^j T_x^* \partial M\big) \oplus \big(\nu^\flat(x) \wedge \Lambda^{j-1} T_x^* \partial M\big), \quad \phi = \phi^{\tan} + \nu^\flat \wedge \phi^{\mathrm{nor}}.$$

The local boundary condition corresponding to the subbundle $E' := \Lambda^* \partial M \subset E_{|\partial M}$ is called the *absolute boundary condition*,

$$B_{\mathrm{abs}} = \{\phi \in H^{1/2}(\partial M, E) \mid \phi^{\mathrm{nor}} = 0\},$$

while $E'' := \nu^\flat \wedge \Lambda^* \partial M \subset E_{|\partial M}$ yields the *relative boundary condition*,

$$B_{\mathrm{rel}} = \{\phi \in H^{1/2}(\partial M, E) \mid \phi^{\tan} = 0\}.$$

Both boundary conditions are known to be elliptic in the classical sense of Lopatinski and Shapiro, see, e.g., [Gi, Lemma 4.1.1]. Indeed, for any $\xi \in T^* \partial M$, the symbol $\sigma_D(\xi)$ leaves the subbundles E' and E'' invariant, while $\sigma_D(\nu^\flat)$ interchanges them. Hence $\sigma_A(\xi)$ interchanges E' and E''. By Corollary 4.18, both, the absolute and the relative boundary condition, are local D-elliptic boundary conditions.

Example 4.20 (Boundary Chirality). Let χ be an orthogonal involution of E along ∂M and denote by $E|_{\partial M} = E^+ \oplus E^-$ the orthogonal splitting into the eigenbundles of χ for the eigenvalues ± 1. We say that χ is a *boundary chirality* (with respect to A) if χ anticommutes with A. The associated boundary conditions $B_{\pm\chi} = H^{1/2}(\partial M, E^\pm)$ are D-elliptic, by Corollary 4.18. In fact, $\chi H^{1/2}_{(-\infty,0)}(A) = H^{1/2}_{(0,\infty)}(A)$ since χ anticommutes with A, and hence

$$B_{\pm\chi} = \{\phi \in \ker A \mid \chi\phi = \pm\phi\} \oplus \{\phi \pm \chi\phi \mid \phi \in H^{1/2}_{(-\infty,0)}(A)\}.$$

We have $B_{-\chi} = B_\chi^\perp$ and hence $\sigma_D(\nu^\flat) B_{-\chi}$ is the adjoint of B_χ.

An example of a boundary chirality is $\chi = i\sigma_D(\nu^\flat)$ in the case where D is formally self-adjoint and A has been chosen to anticommute with χ as in Lemma 3.2. This occurs, for instance, if D is a Dirac operator in the sense of Gromov and Lawson and A is the canonical boundary operator for D; see Appendix 1.

There is a refinement which is due to Freed [Fr, §2]: enumerate the connected components of ∂M as $N_1, \dots, N_k$ and associate a sign $\varepsilon_j \in \{-1, 1\}$ to each component N_j. Then

$$\chi\phi := \sum_j i\varepsilon_j \sigma_D(\nu^\flat)\phi_j,$$

where $\phi_j := \phi_j|_{N_j}$ is again a boundary chirality. It has the additional property that it commutes with $i\sigma_D(\nu^\flat)$; compare Lemma 6.8 and Theorem 6.10.

Example 4.21 (Generalized Atiyah–Patodi–Singer Boundary Conditions). Let D be a Dirac-type operator and A an admissible boundary operator. Fix $a \in \mathbb{R}$ and let

$$V_- := L^2_{(-\infty,a)}(A), \quad V_+ := L^2_{[a,\infty)}(A), \quad W_- = W_+ := \{0\}, \quad \text{and} \quad g = 0.$$

Then the D-elliptic boundary condition

$$B(a) = H^{1/2}_{(-\infty,a)}(A).$$

is known as a *generalized Atiyah–Patodi–Singer boundary condition*. The (non-generalized) Atiyah–Patodi–Singer boundary condition as studied in [APS] is the special case $a = 0$. Generalized APS boundary conditions are not local. However, they are still pseudo-local, by [APS, p. 48] together with [Se] or by [BW, Proposition 14.2].

Example 4.22 (Modified Atiyah–Patodi–Singer Boundary Conditions). The modified APS boundary condition, introduced in [HMR], is given by

$$B_{\mathrm{mAPS}} = \{\phi \in H^{1/2}(\partial M, E) \mid \phi + \sigma_D(\nu^\flat)\phi \in H^{1/2}_{(-\infty,0)}(A)\}.$$

It requires that the spectral parts $\phi = \phi_{(-\infty,0)} + \phi_0 + \phi_{(0,\infty)}$ of $\phi \in B_{\mathrm{mAPS}}$ satisfy

$$\phi_{(0,\infty)} = -\sigma_D(\nu^\flat)\phi_{(-\infty,0)} \quad \text{and} \quad \phi_0 = -\sigma_D(\nu^\flat)\phi_0.$$

Since $\sigma_D(\nu^\flat)^2 = -1$, we get $\phi_0 = 0$. Thus B_{mAPS} is D-elliptic with the choices

$$V_- = L^2_{(-\infty,0)}(A),\; V_+ = L^2_{(0,\infty)}(A),\; W_- = \ker(A),\; W_+ = \{0\}, \text{ and } g = -\sigma_D(\nu^\flat).$$

Example 4.23 (Transmission Conditions). Let M be a complete Riemannian manifold. For the sake of simplicity, assume that the boundary of M is empty, even though this is not really necessary. Let $N \subset M$ be a compact hypersurface with trivial normal bundle. Cut M along N to obtain a Riemannian manifold M' with compact boundary. The boundary $\partial M'$ consists of two copies N_1 and N_2 of N. We may write $M' = (M \setminus N) \sqcup N_1 \sqcup N_2$ (Fig. 1).

Let $E, F \to M$ be Hermitian vector bundles and D be a Dirac-type operator from E to F. We get induced bundles $E' \to M'$ and $F' \to M'$ and a Dirac-type operator D' from E' to F'. For $\Phi \in H^1_{\mathrm{loc}}(M, E)$, we get $\Phi' \in H^1_{\mathrm{loc}}(M', E')$ such that $\Phi'|_{N_1} = \Phi'|_{N_2}$. We use this as a boundary condition for D' on M'. We set

$$B := \left\{(\phi, \phi) \in H^{1/2}(N_1, E) \oplus H^{1/2}(N_2, E) \mid \phi \in H^{1/2}(N, E)\right\},$$

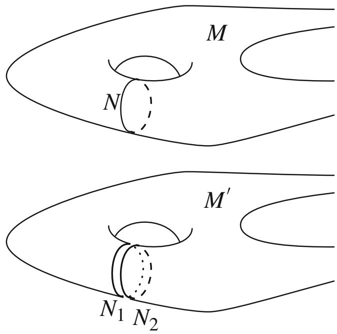


Fig. 1 Cutting M along the hypersurface N

where we identify

$$H^{1/2}(N_1, E) = H^{1/2}(N_2, E) = H^{1/2}(N, E).$$

Let $A = A_0 \oplus -A_0$ be an adapted boundary operator for D'. Here A_0 is a self-adjoint Dirac-type operator on $C^\infty(N, E) = C^\infty(N_1, E')$ and similarly $-A_0$ on $C^\infty(N, E) = C^\infty(N_2, E')$. The sign is due to the opposite relative orientations of N_1 and N_2 in M'.

To see that B is a D'-elliptic boundary condition, put

$$\begin{aligned}
V_+ &:= L^2_{(0,\infty)}(A_0 \oplus -A_0) = L^2_{(0,\infty)}(A_0) \oplus L^2_{(-\infty,0)}(A_0),\\
V_- &:= L^2_{(-\infty,0)}(A_0 \oplus -A_0) = L^2_{(-\infty,0)}(A_0) \oplus L^2_{(0,\infty)}(A_0),\\
W_+ &:= \{(\phi, \phi) \in \ker(A_0) \oplus \ker(A_0)\},\\
W_- &:= \{(\phi, -\phi) \in \ker(A_0) \oplus \ker(A_0)\},
\end{aligned}$$

and

$$g : V_-^{1/2} \to V_+^{1/2}, \quad g = \begin{pmatrix} 0 & \mathrm{id} \\ \mathrm{id} & 0 \end{pmatrix}.$$

With these choices B is of the form required in Definition 4.7. We call these boundary conditions *transmission conditions*. Transmission conditions are not pseudo-local.

If M has a nonempty boundary and N is disjoint from ∂M, let us assume that we are given a D-elliptic boundary condition for ∂M. Then the same discussion applies if one keeps the boundary condition on ∂M and extends B to $\partial M' = \partial M \sqcup N_1 \sqcup N_2$ accordingly.

5 Spectral Theory

Throughout this section we assume the Standard Setup 4.1.

5.1 Coercivity at Infinity

For spectral and index theory we will also need boundary conditions at infinity if M is noncompact. Such conditions go under the name coercivity at infinity.

Definitions 5.1 *For $\kappa > 0$, we say that D is κ-coercive at infinity if there is a compact subset $K \subset M$ such that*

$$\kappa \|\Phi\|_{L^2(M)} \leq \|D\Phi\|_{L^2(M)},$$

for all smooth sections Φ of E with compact support in $M \setminus K$. If D is κ-coercive at infinity for some $\kappa > 0$, then we call D coercive at infinity.

Boundary conditions are irrelevant for coercivity at infinity because the compact set K can always be chosen such that it contains a neighborhood of ∂M.

Examples 5.2. (1) If M is compact, then D is κ-coercive at infinity, for any $\kappa > 0$. Simply choose $K = M$.

(2) If D is formally self-adjoint and, outside a compact subset $K \subset M$, all eigenvalues of the endomorphism $\mathscr{K}$ in the Weitzenböck formula (9) are bounded below by a constant $\kappa > 0$, then we have, for all $\Phi \in C^\infty_{cc}(M, E)$ with support disjoint from K,

$$\|D\Phi\|^2_{L^2(M)} = \|\nabla\Phi\|^2_{L^2(M)} + (\mathscr{K}\Phi, \Phi)_{L^2(M)} \geq \kappa \|\Phi\|^2_{L^2(M)}.$$

Hence D is $\sqrt{\kappa}$-coercive at infinity in this case.

(3) Let $M = S^n \times [0, \infty)$ endowed with the product metric $g_0 + dt^2$, where g_0 is the standard Riemannian metric of the unit sphere and t is the standard coordinate on $[0, \infty)$. Consider the usual Dirac operator D acting on spinors, and denote by ∇ the Levi-Civita connection on the spinor bundle. The Lichnerowicz formula gives

$$D^2 = \nabla^*\nabla + R/4,$$

where $R = n(n-1)$ is the scalar curvature of M (and S^n). It follows that D is $\sqrt{n(n-1)}/2$-coercive at infinity.

(4) Consider the same manifold $M = S^n \times [0, \infty)$, but now equipped with the warped metric $e^{-2t} g_0 + dt^2$. The scalar curvature is easily computed to be

$$R = R(t) = n(n-1)e^{2t} - n(n+1) \to \infty.$$

It follows that this time the Dirac operator D is κ-coercive at infinity, for any $\kappa > 0$.

Theorem 5.3 ([BB, Theorem 8.5]). *Assume the Standard Setup 4.1. Then the following are equivalent:*

(i) D is coercive at infinity;

(ii) $D_{B,\max} : \operatorname{dom} D_{B,\max} \to L^2(M,F)$ *has finite dimensional kernel and closed image for* some *D-elliptic boundary condition B;*

(iii) $D_{B,\max} : \operatorname{dom} D_{B,\max} \to L^2(M,F)$ *has finite dimensional kernel and closed image for* all *D-elliptic boundary conditions B.*

In particular, D and D^* *are coercive at infinity if and only if* $D_{B,\max}$ *and* $D^*_{B^{\mathrm{ad}},\max}$ *are Fredholm operators for some/all D-elliptic boundary conditions B.* ■

Extending the notion of Fredholm operator, we say that a closed operator T between Banach spaces X and Y is a *left-* or *right-Fredholm operator* if the image of T is closed and, respectively, the kernel or the cokernel of T is of finite dimension. We say that T is a *semi-Fredholm operator* if it is a left- or right-Fredholm operator, compare [Ka, Sect. IV.5.1]. In this terminology, Theorem 5.3 says that $D_{B,\max}$ is a left-Fredholm operator for some/all B if and only if D is coercive at infinity. For more on this topic, see [Ka, IV.4 and IV.5], [BBC, Appendix A], and [BB, Appendix A].

In the case $X = Y$, we get corresponding *essential* parts of the spectrum of T, compare [Ka, Sect. IV.5.6] (together with footnotes). We let

$$\operatorname{spec}_{\mathrm{ess}} T \subset \operatorname{spec}_{\mathrm{nlf}} T \subset \operatorname{spec}_{\mathrm{nf}} T \subset \operatorname{spec} T$$

be the set of $\lambda \in \mathbb{C}$ such that $T - \lambda$ is not a semi-Fredholm operator, not a left-Fredholm operator, not a Fredholm operator, and not an isomorphism from $\operatorname{dom} T$ to X, respectively, where *ess* stands for *essential*. In the case where X is a Hilbert space and where T is self-adjoint, $\ker T = (\operatorname{im} T)^\perp$ and $\operatorname{spec} T \subset \mathbb{R}$ so that, in particular, $\operatorname{spec}_{\mathrm{ess}} T = \operatorname{spec}_{\mathrm{nf}} T$. Moreover, in this case, $\operatorname{spec} T \setminus \operatorname{spec}_{\mathrm{ess}} T$ consists of eigenvalues with finite multiplicities, see Remark 1.11 in [Ka, Sect. X.1.2].

Corollary 5.4. *Assume the Standard Setup 4.1 and* $E = F$*. Let* $B \subset H^{1/2}(\partial M, E)$ *be a D-elliptic boundary condition. Let* $\kappa > 0$ *and assume that D is* κ*-coercive at infinity. Then*

$$\{z \in \mathbb{C} \mid |z| < \kappa\} \cap \operatorname{spec}_{nlf} D_{B,\max} = \emptyset.$$

If D and D^* *are* κ*-coercive at infinity, then*

$$\{z \in \mathbb{C} \mid |z| < \kappa\} \cap \operatorname{spec}_{nf} D_{B,\max} = \emptyset.$$

Proof. For any $z \in \mathbb{C}$, the operators $D - z$ and $(D - z)^* = D^* - \bar{z}$ are of Dirac type such that $(D - z)_{\max} = D_{\max} - z$ and $(D^* - \bar{z})_{\max} = D^*_{\max} - \bar{z}$. Moreover, B is a $(D - z)$-elliptic and B^{ad} a $(D^* - \bar{z})$-elliptic boundary condition, one the adjoint of the other. By the triangle inequality, if D is κ-coercive and $|z| < \kappa$, then $D - z$ is $(\kappa - |z|)$-coercive, and similarly for $D^* - \bar{z}$. Thus Theorem 5.3 applies. □

Corollary 5.5. *Assume the Standard Setup 4.1, that $E = F$, and that D is formally self-adjoint. Let $B \subset H^{1/2}(\partial M, E)$ be a self-adjoint D-elliptic boundary condition. If D is κ-coercive at infinity for* some $\kappa > 0$, *then $D_{B,\max}$ is self-adjoint with*

$$(-\kappa, \kappa) \cap \operatorname{spec}_{\mathrm{ess}} D_{B,\max} = \emptyset. \qquad \square$$

Corollary 5.6. *Assume the Standard Setup 4.1, that $E = F$, and that D is formally self-adjoint. Let $B \subset H^{1/2}(\partial M, E)$ be a self-adjoint D-elliptic boundary condition. If D is κ-coercive at infinity for* all $\kappa > 0$, *then $D_{B,\max}$ is self-adjoint with*

$$\operatorname{spec}_{\mathrm{ess}} D_{B,\max} = \emptyset.$$

In particular, the eigenspaces of D are finite dimensional, pairwise L^2-orthogonal, and their sum spans $L^2(M, E)$ in the sense of Hilbert spaces. Moreover, eigensections of D are smooth on M (up to the boundary). $\square$

Remark 5.7. If M is compact, then D is κ-coercive at infinity for all $\kappa > 0$. Hence Corollary 5.6 applies if M is compact with boundary. On the other hand, the resolvent of $D_{B,\max}$ is compact in this case so that the decomposition of $L^2(M, E)$ into finite dimensional eigenspaces is also clear from this perspective.

5.2 Coercivity with Respect to a Boundary Condition

Now we discuss spectral gaps of D about 0. We get interesting results for Dirac operators in the sense of Gromov and Lawson, see Appendix 1.

Definition 5.8. For $\kappa > 0$, we say that D is *κ-coercive with respect to a boundary condition B* if

$$\kappa \|\Phi\|_{L^2(M)} \le \|D\Phi\|_{L^2(M)},$$

for all $\Phi \in C_c^\infty(M, E; B)$.

In contrast to coercivity at infinity, the boundary condition B is now crucial for the concept of coercivity.

Corollary 5.9. *Assume the Standard Setup 4.1, that $E = F$, and that D is formally self-adjoint. Let $B \subset H^{1/2}(\partial M, E)$ be a self-adjoint D-elliptic boundary condition. If D is κ-coercive with respect to B, for $\kappa > 0$, then $D_{B,\max}$ is self-adjoint with*

$$(-\kappa, \kappa) \cap \operatorname{spec} D_{B,\max} = \emptyset.$$

$\square$

Theorem 5.10. *Assume the Standard Setup 4.1 with $E = F$ and that*

$\diamond$ *D is a Dirac operator in the sense of Gromov and Lawson;*

$\diamond$ *B is a D-elliptic boundary condition;*

$\diamond$ *the canonical boundary operator $A : C^\infty(\partial M, E) \to C^\infty(\partial M, E)$ for D satisfies*

$$\big((A - \tfrac{n-1}{2}H)\phi, \phi\big) \le 0$$

for all $\phi \in B$, where H is the mean curvature H along ∂M with respect to the interior unit normal vector field ν;

$\diamond$ *the endomorphism field $\mathscr{K}$ in the Weitzenböck formula (9) satisfies $\mathscr{K} \ge \kappa > 0$. Then D is $\sqrt{\frac{n\kappa}{n-1}}$-coercive with respect to B. In particular, if B is self-adjoint, then*

$$\Big(-\sqrt{\tfrac{n\kappa}{n-1}}, \sqrt{\tfrac{n\kappa}{n-1}}\Big) \cap \operatorname{spec} D_{B,\max} = \emptyset.$$

Proof. For any $\Phi \in C_c^\infty(M, E; B)$ we have by (27) and (28), again writing $\phi = \Phi|_{\partial M}$,

$$\tfrac{n-1}{n}\|D\Phi\|^2 \ge \int_M \langle \mathscr{K}\Phi, \Phi\rangle \mathrm{dV} - \int_{\partial M} (A - \tfrac{n-1}{2}H)|\phi|^2 \,\mathrm{dS} \ge \kappa \|\Phi\|^2. \tag{21}$$

This proves $\sqrt{\frac{n\kappa}{n-1}}$-coerciveness with respect to B. The statement on the spectrum now follows from Corollary 5.9. □

Here are some boundary conditions for which Theorem 5.10 applies:

Example 5.11. Let χ be a boundary chirality with associated D-elliptic boundary condition $B_{\pm\chi} = H^{1/2}(\partial M, E^\pm)$ as in Example 4.20. For $\phi, \psi \in B_\chi$, we have

$$(A\phi, \psi) = (A\chi\phi, \psi) = -(\chi A\phi, \psi) = -(A\phi, \chi\psi) = -(A\phi, \psi).$$

Hence $(A\phi, \psi) = 0$, for all $\phi, \psi \in B_\chi$, and similarly for $B_{-\chi}$. If χ anticommutes with $\sigma_D(\nu^\flat)$, then B_χ and $B_{-\chi}$ are self-adjoint boundary conditions[3]. Hence Theorem 5.10 applies if $H \ge 0$. In the case of the classical Dirac operator D acting on spinors, this yields the eigenvalue estimate in [HMR, Theorem 3].

Example 5.12. The Atiyah–Patodi–Singer boundary condition

$$B_{\mathrm{APS}} = H^{1/2}_{(-\infty,0)}(A)$$

is D-elliptic with adjoint boundary condition

$$B_{\mathrm{APS}}{}^{\mathrm{ad}} = B_{\mathrm{APS}} \oplus \ker A.$$

[3] If χ commutes with $\sigma_D(\nu^\flat)$, then B_χ and $B_{-\chi}$ are adjoint to each other.

Hence $D_{B_{\mathrm{APS}},\max}$ is symmetric. If $\ker A$ is trivial, then $D_{B_{APS}}$ is self-adjoint and $\operatorname{spec} D_{B_{APS}} \subset \mathbb{R}$. By definition of B_{APS}, we have $(A\phi, \phi) \leq -\mu_1 \|\phi\|^2_{L^2(\partial M)}$ for all $\phi \in B_{\mathrm{APS}}$ where $-\mu_1$ is the largest negative eigenvalue of A. Hence Theorem 5.10 applies if $H \geq -\frac{2}{n-1}\mu_1$.

In the case of the classical Dirac operator D acting on spinors, this yields the eigenvalue estimate for the APS boundary condition in [HMR, Theorem 2]. Note that the assumption $\ker A = 0$ is missing in Theorem 2 of [HMR]. In fact, if $\ker A$ is nontrivial, then $D_{B_{APS}}$ is not self-adjoint and $\operatorname{spec} D_{B_{APS}} = \mathbb{C}$, compare [Ka, Sect. V.3.4].

If we can choose a subspace $L \subset \ker A$ as in Theorem 4.12.2, then $B = H^{1/2}_{(-\infty,0)}(A) \oplus L$ is a self-adjoint D-elliptic boundary condition. We have $(A\phi, \phi) \leq 0$ for all $\phi \in B$ and Theorem 5.10 applies if $H \geq 0$.

Example 5.13. The modified APS boundary condition

$$B_{\mathrm{mAPS}} = \{\phi \in H^{1/2}(\partial M, E) \mid \phi + \sigma_D(\nu^\flat)\phi \in H^{1/2}_{(-\infty,0)}(A)\}$$

as in Example 4.22 is D-elliptic with adjoint condition

$$\begin{aligned} B^{\mathrm{ad}}_{\mathrm{mAPS}} &= \{\phi \in H^{1/2}(\partial M, E) \mid \phi_{(0,\infty)} = -\sigma_D(\nu^\flat)\phi_{(-\infty,0)}\} \\ &= B_{\mathrm{mAPS}} \oplus \ker A. \end{aligned}$$

Hence $D_{B_{\mathrm{mAPS}},\max}$ is symmetric. The remaining part of the discussion is as in the previous example, except that we have $(A\phi, \psi) = 0$, for all $\phi, \psi \in B^{\mathrm{ad}}_{\mathrm{mAPS}}$. In particular, Theorem 5.10 applies if $\ker A = 0$ and $H \geq 0$. In the case of the classical Dirac operator D acting on spinors, this yields the eigenvalue estimate in [HMR, Theorem 5]. As in the case of the APS boundary condition, the requirement $\ker A = 0$ needs to be added to the assumptions of Theorem 5 in [HMR].

Next we discuss under which circumstances the "extremal values" $\pm\sqrt{\frac{n\kappa}{n-1}}$ actually belong to the spectrum. For this purpose, we make the following:

Definition 5.14. Let D be a formally self-adjoint Dirac operator in the sense of Gromov and Lawson with associated connection ∇. A section $\Phi \in C^\infty(M, E)$ is called a *D-Killing section* if

$$\nabla_X \Phi = \alpha \cdot \sigma_D(X^\flat)^* \Phi \tag{22}$$

for some constant $\alpha \in \mathbb{R}$ and all $X \in TM$. The constant α is called the *Killing constant* of Φ.

Remarks 5.15. (1) If D is the classical Dirac operator, then spinors satisfying (22) are called Killing spinors. This motivates the terminology.
(2) Equation (22) is overdetermined elliptic. Hence the existence of a nontrivial solution imposes strong restrictions on the underlying geometry. For instance, if a Riemannian spin manifold carries a nontrivial Killing spinor, it must be Einstein [Fri, Theorem B]. See [B] for a classification of manifolds admitting Killing spinors.

(3) Any D-Killing section with Killing constant α is an eigensection of D for the eigenvalue $n\alpha$:

$$D\Phi = \sum_{j=1}^{n} \sigma_D(e_j^\flat)\nabla_{e_j}\Phi = \alpha \cdot \sum_{j=1}^{n} \sigma_D(e_j^\flat)\sigma_D(e_j^\flat)^*\Phi = n\alpha\Phi.$$

(4) Any D-Killing section satisfies the twistor equation (29):

$$\nabla_X\Phi = \alpha \cdot \sigma_D(X^\flat)^*\Phi = \frac{1}{n} \cdot \sigma_D(X^\flat)^* D\Phi.$$

(5) Since $\sigma_D(X^\flat)$ is skewhermitian, the connection $\hat{\nabla}_X = \nabla_X - \alpha \cdot \sigma_D(X^\flat)^*$ is also a metric connection. Since D-Killing sections are precisely $\hat{\nabla}$-parallel sections, we conclude that any D-Killing section Φ has constant length $|\Phi|$.

Theorem 5.16. *In addition to the assumptions in Theorem 5.10 assume that M is compact and that the boundary condition B is self-adjoint.*

Then $\sqrt{\frac{n\kappa}{n-1}} \in \mathrm{spec}(D)$ or $-\sqrt{\frac{n\kappa}{n-1}} \in \mathrm{spec}(D)$ if and only if there is a nontrivial D-Killing section Φ with $\phi = \Phi|_{\partial M} \in B$ and Killing constant $\sqrt{\frac{\kappa}{n(n-1)}}$ or $-\sqrt{\frac{\kappa}{n(n-1)}}$, respectively.

Proof. Let $\sqrt{\frac{n\kappa}{n-1}} \in \mathrm{spec}(D)$, the case $-\sqrt{\frac{n\kappa}{n-1}} \in \mathrm{spec}(D)$ being treated similarly. Since M is compact, the spectral value $\sqrt{\frac{n\kappa}{n-1}}$ must be an eigenvalue by Corollary 5.5. Let Φ be an eigensection of D for the eigenvalue $\sqrt{\frac{n\kappa}{n-1}}$ satisfying the boundary condition. Then we must have equality everywhere in the chain of inequalities (21). In particular, Φ must solve the twistor equation (29). Hence

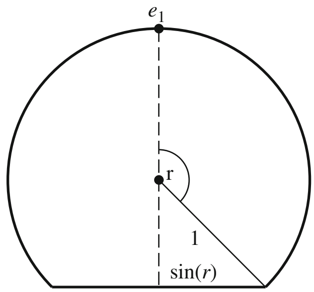
$$\nabla_X\Phi = \tfrac{1}{n}\sigma_D(X^\flat)^* D\Phi = \sqrt{\tfrac{\kappa}{n(n-1)}}\sigma_D(X^\flat)^*\Phi.$$

Conversely, if Φ is a D-Killing section with Killing constant $\sqrt{\frac{\kappa}{n(n-1)}}$, then Φ is an eigensection of D for the eigenvalue $\sqrt{\frac{n\kappa}{n-1}}$, by Remark 5.15.3. □

Example 5.17. Let M be the closed geodesic ball of radius $r \in (0, \pi)$ about e_1 in the unit sphere S^n. The sectional curvature of M is identically equal to 1, its scalar curvature to $n(n-1)$. The boundary ∂M is a round sphere of radius $\sin(r)$ (Fig. 2). Its mean curvature with respect to the interior unit normal is given by $H = \cot(r)$.

We consider the classical Dirac operator acting on spinors. The restriction of the spinor bundle to the boundary yields the spinor bundle of the boundary if n is odd and the sum of two copies of the spinor bundle of the boundary if n is even. Accordingly, the canonical boundary operator is just the classical Dirac operator of the boundary if n is odd and the direct sum of it and its negative if n is even. The kernel of the boundary operator is trivial.

Fig. 2 Geodesic ball of radius r in the unit sphere



Theorem 5.10 applies with all the boundary conditions described in Examples 5.11–5.13 if $r \leq \frac{\pi}{2}$ because then $H \geq 0$. Therefore the spectrum of the Dirac operator on M subject to any of these boundary conditions does not intersect $(-\frac{n}{2}, \frac{n}{2})$. The largest negative Dirac eigenvalue of the boundary is given by $-\mu_1 = -\frac{n-1}{2\sin(r)}$. Since we have

$$\tfrac{n-1}{2}H = \tfrac{n-1}{2}\cot(r) \geq -\tfrac{n-1}{2}\sin(r) = -\mu_1,$$

Theorem 5.10 applies in the case of APS boundary conditions (Example 5.12) for all $r \in (0, \pi)$.

The sphere S^n and hence M do possess nontrivial Killing spinors for both Killing constants $\pm\frac{1}{2}$. The restriction of such a Killing spinor to ∂M never satisfies the APS boundary conditions. Thus the equality case in Theorem 5.10 does not occur and $\pm\frac{n}{2}$ cannot lie in the spectrum of D on M subject to APS conditions. Hence, under APS boundary conditions and for any $r \in (0, \pi)$, the spectrum of D on M does not intersect $[-\frac{n}{2}, \frac{n}{2}]$.

The modified APS boundary conditions are satisfied by the restrictions of the Killing spinors only if $r = \frac{\pi}{2}$. In this case, $\frac{n}{2}$ is an eigenvalue of D on M.

6 Index Theory

Throughout this section, assume the Standard Setup 4.1. In Theorem 5.3 we have seen that $D_{B,\max} : \operatorname{dom} D_{B,\max} \to L^2(M, F)$ is a Fredholm operator for any D-elliptic boundary condition provided D and D^* are coercive at infinity. This is the case if M is compact, for instance. The index is the number

$$\operatorname{ind} D_{B,\max} = \dim\ker D_{B,\max} - \dim\ker D^*_{B^{\mathrm{ad}},\max} \in \mathbb{Z}.$$

If B is a D-elliptic boundary condition, then, by Theorems 4.2.4 and 4.4, $D_{B,\max}$ has domain

$$\operatorname{dom} D_{B,\max} = \{\Phi \in \operatorname{dom} D_{\max} \mid \mathscr{R}\Phi \in B\} \subset H^1_{\mathrm{loc}}(M, E).$$

Since $\operatorname{dom} D_{B,\max}$ is contained in $H^1_{\mathrm{loc}}(M, E)$, we will briefly write D_B instead of $D_{B,\max}$.

6.1 Fredholm Property and Index Formulas

As a direct consequence of Theorem 5.3 we get

Corollary 6.1 ([BB, Corollary 8.7]). *Assume the Standard Setup 4.1 and that D and D^* are coercive at infinity. Let B be a D-elliptic boundary condition and let $\check{C}$ be a closed complement of B in $\check{H}(A)$. Let $\check{P} : \check{H}(A) \to \check{H}(A)$ be the projection with kernel B and image $\check{C}$. Then*

$$\check{L} : \operatorname{dom} D_{\max} \to L^2(M, F) \oplus \check{C}, \quad \check{L}\Phi = (D_{\max}\Phi, \check{P}\mathscr{R}\Phi),$$

is a Fredholm operator with the same index as D_B. □

Corollary 6.2 ([BB, Corollary 8.8]). *Assume the Standard Setup 4.1 and that D and D^* are coercive at infinity. Let $B_1 \subset B_2 \subset H^{1/2}(\partial M, E)$ be D-elliptic boundary conditions for D. Then* $\dim(B_2/B_1)$ *is finite and*

$$\operatorname{ind}(D_{B_2}) = \operatorname{ind}(D_{B_1}) + \dim(B_2/B_1). \qquad \blacksquare$$

Example 6.3. For the generalized Atiyah–Patodi–Singer boundary conditions as in Example 4.21 and $a < b$, we have

$$\operatorname{ind} D_{B(b)} = \operatorname{ind} D_{B(a)} + \dim L^2_{[a,b)}(A).$$

The following result says that index computations for D-elliptic boundary conditions can be reduced to the case of generalized Atiyah–Patodi–Singer boundary conditions.

Theorem 6.4 ([BB, Theorem 8.14]). *Assume the Standard Setup 4.1 and that D and D^* are coercive at infinity. Let $B \subset H^{1/2}(\partial M, E)$ be a D-elliptic boundary condition. Then we have, in the representation of B as in Remark 4.8.2,*

$$\operatorname{ind} D_B = \operatorname{ind} D_{B(a)} + \dim W_+ - \dim W_-.$$

Sketch of Proof. Replacing g by sg, $s \in [0, 1]$, yields a continuous 1-parameter family of D-elliptic boundary conditions. One can show that the index stays constant under such a deformation of boundary conditions. Therefore, we can assume without loss of generality that $g = 0$, i.e., $B = W_+ \oplus V_-^{1/2}$. Consider one further boundary condition,

$$B' := W_- \oplus W_+ \oplus V_-^{1/2} = H^{1/2}_{(-\infty,a)}(A) \oplus W_+ = B(a) \oplus W_+.$$

Applying Corollary 6.2 twice we conclude

$$\mathrm{ind}(D_B) = \mathrm{ind}(D_{B'}) - \dim W_- = \mathrm{ind}(D_{B(a)}) + \dim W_+ - \dim W_-. \qquad \square$$

6.2 Relative Index Theory

Assume the Standard Setup 4.1 throughout the section. For convenience assume also that M is connected and that $\partial M = \emptyset$. For what follows, compare Example 4.23. Let N be a closed and two-sided hypersurface in M. Cut M along N to obtain a manifold M', possibly connected, whose boundary $\partial M'$ consists of two disjoint copies N_1 and N_2 of N, see Fig. 1 on page 61. There are natural pull-backs E', F', and D' of E, F, and D from M to M'. Choose an adapted operator A for D' along N_1. Then $-A$ is an adapted operator for D' along N_2 and will be used in what follows.

Theorem 6.5 (Splitting Theorem, [BB, Theorem 8.17]). *For M, M', and notation as above, D and D^* are coercive at infinity if and only if D' and $(D')^*$ are coercive at infinity. In this case, D and $D'_{B_1\oplus B_2}$ are Fredholm operators with*

$$\mathrm{ind} D = \mathrm{ind} D'_{B_1\oplus B_2},$$

where $B_1 = B(a) = H^{1/2}_{(-\infty,a)}(A)$ and $B_2 = H^{1/2}_{[a,\infty)}(A)$, considered as boundary conditions along N_1 and N_2, respectively. More generally, we may choose any D-elliptic boundary condition $B_1 \subset H^{1/2}(N,E)$ and its L^2-orthogonal complement $B_2 \subset H^{1/2}(N,E)$. ■

Let M_1 and M_2 be complete Riemannian manifolds without boundary and

$$D_i : C^\infty(M_i, E_i) \to C^\infty(M_i, F_i)$$

be Dirac-type operators. Let $K_1 \subset M_1$ and $K_2 \subset M_2$ be compact subsets. Then we say that D_1 *outside* K_1 *agrees with* D_2 *outside* K_2 if there are an isometry $f : M_1 \setminus K_1 \to M_2 \setminus K_2$ and smooth fiberwise linear isometries

$$\mathscr{I}_E : E_1|_{M_1\setminus K_1} \to E_2|_{M_2\setminus K_2} \quad \text{and} \quad \mathscr{I}_F : F_1|_{M_1\setminus K_1} \to F_2|_{M_2\setminus K_2}$$

such that

$$\begin{array}{ccc} E_1|_{M_1\setminus K_1} & \xrightarrow{\mathscr{I}_E} & E_2|_{M_2\setminus K_2} \\ \downarrow & & \downarrow \\ M_1\setminus K_1 & \xrightarrow{f} & M_2\setminus K_2 \end{array} \qquad \begin{array}{ccc} F_1|_{M_1\setminus K_1} & \xrightarrow{\mathscr{I}_F} & F_2|_{M_2\setminus K_2} \\ \downarrow & & \downarrow \\ M_1\setminus K_1 & \xrightarrow{f} & M_2\setminus K_2 \end{array}$$

commute and

$$\mathscr{I}_F \circ (D_1\Phi) \circ f^{-1} = D_2(\mathscr{I}_E \circ \Phi \circ f^{-1})$$

for all smooth sections Φ of E_1 over $M_1 \setminus K_1$.

Assume now that D_1 and D_2 agree outside compact domains $K_i \subset M_i$. For $i = 1, 2$, choose a decomposition $M_i = M_i' \cup M_i''$ such that $N_i = M_i' \cap M_i''$ is a compact hypersurface in M_i, K_i is contained in the interior of M_i', $f(M_1'') = M_2''$, and $f(N_1) = N_2$. Denote the restriction of D_i to M_i' by D_i'. The following result is a general version of the Φ-relative index theorem of Gromov and Lawson [GL, Theorem 4.35].

Theorem 6.6 ([BB, Theorem 1.21]). *Under the above assumptions, let $B_1 \subset H^{1/2}(N_1, E_1)$ and $B_2 \subset H^{1/2}(N_2, E_2)$ be D_i-elliptic boundary conditions which correspond to each other under the identifications given by f and $\mathscr{I}_E$ as above. Assume that D_1 and D_2 and their formal adjoints are coercive at infinity.*

Then D_1, D_2, D'_{1,B_1}, and D'_{2,B_2} are Fredholm operators such that

$$\operatorname{ind} D_1 - \operatorname{ind} D_2 = \operatorname{ind} D'_{1,B_1} - \operatorname{ind} D'_{2,B_2} = \int_{K_1} \alpha_{D_1} - \int_{K_2} \alpha_{D_2},$$

where α_{D_1} and α_{D_2} are the index densities associated with D_1 and D_2. ■

Remark 6.7. In Theorem 6.6, it is also possible to deal with the situation that M_1 and M_2 have compact boundary and elliptic boundary conditions B_1 and B_2 along their boundaries are given. One then chooses the hypersurface $N = N_i$ such that it does not intersect the boundary of M_i and such that the boundary of M_i is contained in M_i'. The same arguments as above yield

$$\operatorname{ind} D_{1,B_1} - \operatorname{ind} D_{2,B_2} = \operatorname{ind} D'_{1,B_1 \oplus B_1'} - \operatorname{ind} D'_{2,B_2 \oplus B_2'},$$

where B_1' and B_2' are elliptic boundary conditions along N_1 and N_2 which correspond to each other under the identifications given by f and $\mathscr{I}_E$ as further up. A similar remark applies to Theorem 6.5.

6.3 Boundary Chiralities and Index

Lemma 6.8. *Assume the Standard Setup 4.1 and that M is connected. Let D be formally self-adjoint and let A anticommute with $\sigma_D(\nu^\flat)$. Let χ be a boundary chirality as in Example 4.20 which commutes with $\sigma_D(\nu^\flat)$. Let $E = E^+ \oplus E^-$ be the orthogonal splitting into the eigenbundles of χ for the eigenvalues ± 1, and write*

$$A = \begin{pmatrix} 0 & A^- \\ A^+ & 0 \end{pmatrix}$$

with respect to this splitting. Then, if D is coercive at infinity,

$$\mathrm{ind} D_{B_\chi} = \tfrac{1}{2}\mathrm{ind} A^+ = -\tfrac{1}{2}\mathrm{ind} A^-,$$

where $B_\chi = H^{1/2}(\partial M, E^+)$ is as in Example 4.20.

Proof. Let $B_\pm = \ker A \oplus \{\phi \pm \chi\phi \mid \phi \in H^{1/2}_{(-\infty)}(A)$. Then $B_\pm$ is a D-elliptic boundary condition and, by Theorem 6.4,

$$\mathrm{ind} D_{B_\pm} = \mathrm{ind} D_{B_{\mathrm{APS}}} + \dim \ker A.$$

By Corollary 6.2, we have

$$\mathrm{ind} D_{B_{\pm\chi}} = \mathrm{ind} D_{B_\pm} - \dim \ker A^\mp,$$

where $B_{+\chi} = B_\chi$ and $B_{-\chi} = H^{1/2}(\partial M, E^-)$. Since $B_{-\chi} = B_\chi^\perp$ and $B_{-\chi}$ is invariant under $\sigma_D(\nu^\flat)$, we get that $B_{-\chi}$ is the adjoint of B_χ. In conclusion

$$\begin{aligned} 2\mathrm{ind} D_{B_\chi} &= \mathrm{ind} D_{B_\chi} - \mathrm{ind} D_{B_{-\chi}} \\ &= \mathrm{ind} D_{B_+} - \dim \ker A^- - \mathrm{ind} D_{B_-} + \dim \ker A^+ \\ &= \mathrm{ind} A^+. \end{aligned}$$

□

Theorem 6.9 (Cobordism Theorem, [BB, Theorem 1.22]). *Assume the Standard Setup 4.1 and that M is connected. Let D be formally self-adjoint and let A anticommute with $\sigma_D(\nu^\flat)$. Then $\chi = i\sigma_D(\nu^\flat)$ is a boundary chirality. Moreover, if D is coercive at infinity and with $A^\pm$ as in Theorem 6.8, then*

$$\mathrm{ind} A^+ = \mathrm{ind} A^- = 0.$$

Originally, the cobordism theorem was formulated for compact manifolds M with boundary and showed the cobordism invariance of the index. This played an important role in the original proof of the Atiyah–Singer index theorem, compare, e.g., [Pa, Chap. XVII] and [BW, Chap. 21]. In this case, one can also derive the cobordism invariance from Roe's index theorem for partitioned manifolds [Ro, Hi]. We replace compactness of the bordism by the weaker assumption of coercivity of D.

Sketch of Proof of Theorem 6.9. We show that $\ker D_{B_\chi} = \ker D_{B_{-\chi}} = 0$, then the assertion follows from Lemma 6.8. Let $\Phi \in \ker D_{B^\pm,\max}$. By Theorem 4.2.5, we have

$$\begin{aligned} 0 &= (D_{\max}\Phi, \Phi)_{L^2(M)} - (\Phi, D_{\max}\Phi)_{L^2(M)} \\ &= -(\sigma_D(\nu^\flat)\mathscr{R}\Phi, \mathscr{R}\Phi)_{L^2(\partial M)} \\ &= \pm i\|\mathscr{R}\Phi\|^2_{L^2(\partial M)}, \end{aligned}$$

and hence $\mathscr{R}\Phi = 0$. Now an elementary argument involving the unique continuation for solutions of $D\Phi = 0$ implies $\Phi = 0$. □

As an application of Lemma 6.8 and Theorem 6.9, we generalize Freed's Theorem B from [Fr] as follows:

Theorem 6.10. *Assume the Standard Setup 4.1 and that M is connected. Let D be formally self-adjoint and let A anticommute with $\sigma_D(\nu^\flat)$. Let χ be a boundary chirality as in Example 4.20 which commutes with $\sigma_D(\nu^\flat)$. Let $E = E^{++} \oplus E^{+-} \oplus E^{-+} \oplus E^{--}$ be the orthogonal splitting into the simultaneous eigenbundles of $i\sigma_D(\nu^\flat)$ and χ for the eigenvalues ± 1.*

Then A maps E^{++} to E^{--} and E^{+-} to E^{-+} and conversely. Moreover, with the corresponding notation for the restrictions of A, we have, if D is coercive at infinity,

$$\mathrm{ind} D_{B_\chi} = \mathrm{ind} A^{++} = -\mathrm{ind} A^{--}.$$

Proof. By Theorem 6.9, we have

$$\mathrm{ind} A^{++} + \mathrm{ind} A^{+-} = \mathrm{ind} A^{--} + \mathrm{ind} A^{-+} = 0.$$

On the other hand, A^{--} is adjoint to A^{++}, hence Lemma 6.8 gives

$$2\mathrm{ind} D_{B_\chi} = \mathrm{ind} A^{++} + \mathrm{ind} A^{-+} = \mathrm{ind} A^{++} - \mathrm{ind} A^{--} = 2\mathrm{ind} A^{++}. \qquad □$$

Appendix 1: Dirac Operators in the Sense of Gromov and Lawson

Here we discuss an important subclass of Dirac-type operators. Note that the connection in Corollary 2.4 is not metric, in general.

Definition A.1. A formally self-adjoint operator $D : C^\infty(M,E) \to C^\infty(M,E)$ of Dirac type is called a *Dirac operator in the sense of Gromov and Lawson* if there exists a metric connection ∇ on E such that

(1) $D = \sum_j \sigma_D(e_j^*) \circ \nabla_{e_j}$, for any local orthonormal tangent frame $(e_1, \dots, e_n)$;

(2) the principal symbol σ_D of D is parallel with respect to ∇ and to the Levi-Civita connection.

This is equivalent to the definition of *generalized Dirac operators* in [GL, Sect. 1] or to *Dirac operators on Dirac bundles* in [LM, Chap. II, § 5].

Remark A.2. For a Dirac operator in the sense of Gromov and Lawson, the connection ∇ in Definition A.1 and the connection in the Weitzenböck formula (9) coincide and are uniquely determined by these properties. We will call ∇ the connection *associated with the Dirac operator D*. Moreover, the endomorphism field $\mathscr{K}$ in the Weitzenböck formula takes the form

$$\mathscr{K} = \frac{1}{2}\sum_{i,j} \sigma_D(e_i^*) \circ \sigma_D(e_j^*) \circ R^\nabla(e_i, e_j)$$

where R^∇ is the curvature tensor of ∇. See [GL, Proposition 2.5] for a proof.

Next, we show how to explicitly construct an adapted operator on the boundary satisfying (13) for a Dirac operator in the sense of Gromov and Lawson. Let ∇ be the associated connection. Along the boundary we define

$$A_0 := \sigma_D(\nu^\flat)^{-1} D - \nabla_\nu = \sigma_D(\nu^\flat)^{-1} \sum_{j=2}^n \sigma_D(e_j^*)\nabla_{e_j}. \tag{23}$$

Here $(e_2, \dots, e_n)$ is any local tangent frame for ∂M. Then A_0 is a first-order differential operator acting on section of $E|_{\partial M} \to \partial M$ with principal symbol $\sigma_{A_0}(\xi) = \sigma_D(\nu^\flat)^{-1}\sigma_D(\xi)$ as required for an adapted boundary operator. From the Weitzenböck formula (9) we get, using Proposition 2.1 twice, once for D and once for ∇, for all $\Phi, \Psi \in C_c^\infty(M,E)$:

$$\begin{aligned} 0 &= \int_M \big(\langle D^2\Phi, \Psi\rangle - \langle \nabla^*\nabla\Phi, \Psi\rangle - \langle \mathscr{K}\Phi, \Psi\rangle\big)\mathrm{dV} \\ &= \int_M \big(\langle D\Phi, D\Psi\rangle - \langle \nabla\Phi, \nabla\Psi\rangle - \langle \mathscr{K}\Phi, \Psi\rangle\big)\mathrm{dV} \end{aligned}$$

$$+ \int_{\partial M} \big(- \langle \sigma_D(\nu^\flat) D\Phi, \Psi \rangle + \langle \sigma_{\nabla^*}(\nu^\flat) \nabla\Phi, \Psi \rangle \big) \, \mathrm{dS} \, . \tag{24}$$

For the boundary contribution we have

$$\begin{aligned} -\langle \sigma_D(\nu^\flat) D\Phi, \Psi \rangle + \langle \sigma_{\nabla^*}(\nu^\flat) \nabla\Phi, \Psi \rangle &= \langle \sigma_D(\nu^\flat)^{-1} D\Phi, \Psi \rangle - \langle \nabla\Phi, \sigma_\nabla(\nu^\flat)\Psi \rangle \\ &= \langle \sigma_D(\nu^\flat)^{-1} D\Phi, \Psi \rangle - \langle \nabla\Phi, \nu^\flat \otimes \Psi \rangle \\ &= \langle \sigma_D(\nu^\flat)^{-1} D\Phi, \Psi \rangle - \langle \nabla_\nu\Phi, \Psi \rangle \\ &= \langle A_0\Phi, \Psi \rangle. \end{aligned} \tag{25}$$

Inserting (25) into (24) we get

$$\int_M \big(\langle D\Phi, D\Psi \rangle - \langle \nabla\Phi, \nabla\Psi \rangle - \langle \mathscr{K}\Phi, \Psi \rangle \big) \mathrm{dV} = - \int_{\partial M} \langle A_0 \phi, \psi \rangle \, \mathrm{dS} \tag{26}$$

where $\phi := \Phi|_{\partial M}$ and $\psi := \Psi|_{\partial M}$. Since the left-hand side of (26) is symmetric in Φ and Ψ, the right-hand side is symmetric as well, hence A_0 is formally self-adjoint. This shows that A_0 is an adapted boundary operator for D.

In general, A_0 does not anticommute with $\sigma_D(\nu^\flat)$ however. We will rectify this by adding a suitable zero-order term. First, let us compute the anticommutator of A_0 and $\sigma_D(\nu^\flat)$:

$$\begin{aligned} \{\sigma_D(\nu^\flat), A_0\}\phi &= \sum_{j=2}^n \sigma_D(e_j^*) \nabla_{e_j}\phi + \sigma_D(\nu^\flat)^{-1} \sum_{j=2}^n \sigma_D(e_j^*) \nabla_{e_j}(\sigma_D(\nu^\flat)\phi) \\ &= \sum_{j=2}^n \Big(\sigma_D(e_j^*) \nabla_{e_j}\phi + \sigma_D(\nu^\flat)^{-1} \sigma_D(e_j^*) \sigma_D(\nu^\flat) \nabla_{e_j}\phi \\ &\quad + \sigma_D(\nu^\flat)^{-1} \sigma_D(e_j^*) \sigma_D(\nabla_{e_j}\nu^\flat)\phi \Big) \\ &= \sigma_D(\nu^\flat)^{-1} \sum_{j=2}^n \sigma_D(e_j^*) \sigma_D(\nabla_{e_j}\nu^\flat)\phi. \end{aligned}$$

Now $\nabla.\nu$ is the negative of the Weingarten map of the boundary with respect to the normal field ν. We choose the orthonormal tangent frame $(e_2, \dots, e_n)$ to consist of eigenvectors of the Weingarten map. The corresponding eigenvalues $\kappa_2, \dots, \kappa_n$ are the *principal curvatures* of ∂M. We get

$$\sum_{j=2}^n \sigma_D(e_j^\flat) \sigma_D(\nabla_{e_j}\nu^\flat) = - \sum_{j=2}^n \sigma_D(e_j^\flat) \sigma_D(\kappa_j e_j^\flat) = \sum_{j=2}^n \kappa_j = (n-1)H,$$

where H is the *mean curvature* of ∂M with respect to ν. Therefore,

$$\{\sigma_D(\nu^\flat), A_0\} = (n-1)H\sigma_D(\nu^\flat)^{-1} = -(n-1)H\sigma_D(\nu^\flat).$$

Since clearly

$$\{\sigma_D(\nu^\flat), (n-1)H\} = 2(n-1)H\sigma_D(\nu^\flat),$$

the operator

$$A := A_0 + \frac{n-1}{2}H = \sigma_D(\nu^\flat)^{-1}D - \nabla_\nu + \frac{n-1}{2}H$$

is an adapted boundary operator for D satisfying (13). From (26) we also have

$$\int_M \big(\langle D\Phi, D\Psi\rangle - \langle \nabla\Phi, \nabla\Psi\rangle - \langle \mathscr{K}\Phi, \Psi\rangle\big)\mathrm{dV} = \int_{\partial M} \langle (\tfrac{n-1}{2}H - A)\phi, \psi\rangle \,\mathrm{dS}\,. \tag{27}$$

Definition A.3. For a Dirac operator D in the sense of Gromov and Lawson as above, we call A the *canonical boundary operator* for D.

Remark A.4. The canonical boundary operator A is again a Dirac operator in the sense of Gromov and Lawson. Namely, define a connection on $E|_{\partial M}$ by

$$\nabla^\partial_X\phi := \nabla_X\phi + \tfrac{1}{2}\sigma_D(\nu^\flat)^{-1}\sigma_D(\nabla_X\nu^\flat)\phi.$$

The Clifford relations (6) show that the term $\sigma_D(\nu^\flat)^{-1}\sigma_D(\nabla_X\nu^\flat) = \sigma_D(\nu^\flat)^*\sigma_D(\nabla_X\nu^\flat)$ is skewhermitian, hence ∇^∂ is a metric connection. By (23), $A_0 = \sum_{j=2}^n \sigma_{A_0}(e_j^*)\circ\nabla_{e_j}$. This, $\sigma_{A_0} = \sigma_A$, and

$$\sum_{j=2}^n \sigma_{A_0}(e_j^*)\sigma_D(\nu^\flat)^*\sigma_D(\nabla_{e_j}\nu^\flat) = \frac{n-1}{2}H$$

show that

$$A = \sum_{j=2}^n \sigma_A(e_j^*)\circ\nabla^\partial_{e_j}.$$

Moreover, a straightforward computation using the Gauss equation for the Levi-Civita connections $\nabla_X\xi = \nabla^\partial_X\xi - \xi(\nabla_X\nu)\nu^\flat$ shows that σ_A is parallel with respect to the boundary connections ∇^∂.

Remark A.5. The triangle inequality and the Cauchy–Schwarz inequality show

$$|D\Phi|^2 = \Big|\sum_{j=1}^n \sigma_D(e_j^\flat)\nabla_{e_j}\Phi\Big|^2 \le \Big(\sum_{j=1}^n |\sigma_D(e_j^\flat)\nabla_{e_j}\Phi|\Big)^2$$

$$\le n\cdot\sum_{j=1}^n |\sigma_D(e_j^\flat)\nabla_{e_j}\Phi|^2 = n\cdot\sum_{j=1}^n \langle\sigma_D(e_j^\flat)^*\sigma_D(e_j^\flat)\nabla_{e_j}\Phi, \nabla_{e_j}\Phi\rangle$$

$$= n\cdot\sum_{j=1}^n |\nabla_{e_j}\Phi|^2 = n\cdot|\nabla\Phi|^2, \tag{28}$$

for any orthonormal tangent frame $(e_1,\dots,e_n)$ and all $\Phi \in C^\infty(M,E)$.

When does equality hold? Equality in the Cauchy–Schwarz inequality implies that all summands $|\sigma_D(e_j^\flat)\nabla_{e_j}\Phi|$ are equal, i.e., $|\sigma_D(e_j^\flat)\nabla_{e_j}\Phi| = |\sigma_D(e_1^\flat)\nabla_{e_1}\Phi|$. Equality in the triangle inequality then implies $\sigma_D(e_j^\flat)\nabla_{e_j}\Phi = \sigma_D(e_1^\flat)\nabla_{e_1}\Phi$ for all j. Thus

$$\sigma_D(e_1^\flat)\nabla_{e_1}\Phi = \frac{1}{n}\sum_{j=1}^n \sigma_D(e_j^\flat)\nabla_{e_j}\Phi = \frac{1}{n}D\Phi,$$

hence $\nabla_{e_1}\Phi = \frac{1}{n}\sigma_D(e_1^\flat)^*D\Phi$. Since e_1 is arbitrary, this shows the *twistor equation*

$$\nabla_X\Phi = \tfrac{1}{n}\sigma_D(X^\flat)^*D\Phi, \tag{29}$$

for all vector fields X on M. Conversely, if Φ solves the twistor equation, one sees directly that equality holds in (28).

Inserting (28) into (27) yields

$$\tfrac{n-1}{n}\int_M |D\Phi|^2\mathrm{dV} \ge \int_M \langle\mathscr{K}\Phi,\Phi\rangle\mathrm{dV} + \int_{\partial M}\langle(\tfrac{n-1}{2}H - A)\phi,\phi\rangle\,\mathrm{dS},$$

for all $\Phi \in C^\infty{}_c(M,E)$, where $\phi := \Phi|_{\partial M}$. Moreover, equality holds if and only if Φ solves the twistor equation (29).

Appendix 2: Proofs of Some Auxiliary Results

In this section we collect the proofs of some of the auxiliary results.

Proof of Proposition 2.3. We start by choosing an arbitrary connection $\bar\nabla$ on E and define

$$\bar D : C^\infty(M,E) \to C^\infty(M,F), \quad \bar D\Phi := \sum\nolimits_j \sigma_D(e_j^*)\bar\nabla_{e_j}\Phi.$$

Then $\bar{D}$ has the same principal symbol as D and, therefore, the difference $S := D - \bar{D}$ is of order 0. In other words, S is a field of homomorphisms from E to F.

Since $\mathscr{A}_D$ is onto, the restriction $\mathscr{A}$ of $\mathscr{A}_D$ to the orthogonal complement of the kernel of $\mathscr{A}_D$ is a fiberwise isomorphism. We put $V := \mathscr{A}^{-1}(S)$ and define a new connection by

$$\nabla := \bar{\nabla} + V.$$

We compute

$$\begin{aligned}\sum_j \sigma_D(e_j^*) \circ \nabla_{e_j} &= \sum_j \sigma_D(e_j^*) \circ \bar{\nabla}_{e_j} + \sum_j \sigma_D(e_j^*) \circ V(e_j)\\ &= \bar{D} + \mathscr{A}_D(V)\\ &= \bar{D} + S = D. \end{aligned}$$

□

Proof of Proposition 3.1. Let $\widetilde{\nabla}$ be any metric connection on E. Then $F := D^*D - \widetilde{\nabla}^*\widetilde{\nabla}$ is formally self-adjoint. Since both, D^*D and $\widetilde{\nabla}^*\widetilde{\nabla}$, have the same principal symbol $-|\xi|^2 \cdot \mathrm{id}$, the operator F is of order at most one. Any other metric connection ∇ on E is of the form $\nabla = \widetilde{\nabla} + B$ where B is a 1-form with values in skewhermitian endomorphisms of E. Hence

$$D^*D = (\nabla - B)^*(\nabla - B) + F = \nabla^*\nabla \underbrace{-\nabla^*B - B^*\nabla + B^*B + F}_{=:\mathscr{K}}.$$

In general, $\mathscr{K}$ is of first order and we need to show that there is a unique B such that $\mathscr{K}$ is of order zero. Since B^*B is of order zero, $\mathscr{K}$ is of order zero if and only if $F - \nabla^*B - B^*\nabla$ is of order zero, i.e., if and only if $\sigma_F(\xi) = \sigma_{\nabla^*B+B^*\nabla}(\xi)$ for all $\xi \in T^*M$. We compute, using a local tangent frame $e_1, \ldots, e_n$,

$$\begin{aligned}\langle \sigma_{\nabla^*B+B^*\nabla}(\xi)\varphi, \psi\rangle &= \big\langle\big(\sigma_{\nabla^*}(\xi) \circ B + B^* \circ \sigma_\nabla(\xi)\big)\varphi, \psi\big\rangle\\ &= -\langle B\varphi, \sigma_\nabla(\xi)\psi\rangle + \langle \sigma_\nabla(\xi)\varphi, B\psi\rangle\\ &= -\langle B\varphi, \xi \otimes \psi\rangle + \langle \xi \otimes \varphi, B\psi\rangle\\ &= -\Big\langle \sum_i e_i^* \otimes B_{e_i}\varphi, \xi \otimes \psi\Big\rangle + \Big\langle \xi \otimes \varphi, \sum_i e_i^* \otimes B_{e_i}\psi\Big\rangle\\ &= -\sum_i \langle e_i^*, \xi\rangle \langle B_{e_i}\varphi, \psi\rangle + \sum_i \langle e_i^*, \xi\rangle\, \langle \varphi, B_{e_i}\psi\rangle\\ &= -\big\langle B_{\xi^\sharp}\varphi, \psi\big\rangle + \big\langle \varphi, B_{\xi^\sharp}\psi\big\rangle\\ &= -2\big\langle B_{\xi^\sharp}\varphi, \psi\big\rangle. \end{aligned}$$

Hence, $\sigma_{\nabla^* B + B^* \nabla}(\xi) = -2B_{\xi^\sharp}$. Thus, $\mathscr{K}$ is of order 0 if and only if

$$B_X = -\tfrac{1}{2}\,\sigma_F(X^b)$$

for all $X \in TM$. Note that $\sigma_F(\xi)$ is indeed skewhermitian because F is formally self-adjoint. □

Proof of Lemma 3.2. Since D is formally self-adjoint and of Dirac type,

$$-\sigma_D(\nu^\flat) = \sigma_D(\nu^\flat)^* = \sigma_D(\nu^\flat)^{-1}, \tag{30}$$

by (1) and (8). Let A_0 be adapted to D along ∂M and $\xi \in T_x^*\partial M$, as usual extended to $T_x^* M$ by $\xi(\nu(x)) = 0$. Then, again using (6) and (11),

$$\begin{aligned}
\sigma_{A_0}(\xi) &+ \sigma_D(\nu(x)^\flat)\sigma_{A_0}(\xi)\sigma_D(\nu(x)^\flat)^* \\
&= \sigma_D(\nu(x)^\flat)^{-1}\sigma_D(\xi) + \sigma_D(\xi)\sigma_D(\nu(x)^\flat)^* \\
&= \sigma_D(\nu(x)^\flat)^*\sigma_D(\xi) + \sigma_D(\xi)^*\sigma_D(\nu(x)^\flat) \\
&= 2\langle \nu(x)^\flat, \xi\rangle \cdot \mathrm{id}_E \\
&= 0.
\end{aligned}$$

Hence $2S := A_0 + \sigma_D(\nu^\flat)A_0\sigma_D(\nu^\flat)^*$ is of order 0, that is, S is a field of endomorphisms of E along ∂M. Since A_0 is formally self-adjoint so is S and, by (30),

$$\sigma_D(\nu^\flat)2S = \sigma_D(\nu^\flat)A_0 + A_0\sigma_D(\nu^\flat) = 2S\sigma_D(\nu^\flat).$$

Hence $A := A_0 - S$ is adapted to D along ∂M and

$$\begin{aligned}
\sigma_D(\nu^\flat)A + A\sigma_D(\nu^\flat) &= \sigma_D(\nu^\flat)A_0 + A_0\sigma_D(\nu^\flat) - \sigma_D(\nu^\flat)S - S\sigma_D(\nu^\flat) \\
&= \sigma_D(\nu^\flat)\big(A_0 - \sigma_D(\nu^\flat)A_0\sigma_D(\nu^\flat) - 2S\big) \\
&= \sigma_D(\nu^\flat)\big(A_0 + \sigma_D(\nu^\flat)A_0\sigma_D(\nu^\flat)^* - 2S\big) \\
&= 0.
\end{aligned}$$

□

References

[AT] T. Ackermann, J. Tolksdorf, The generalized Lichnerowicz formula and analysis of Dirac operators. J. Reine Angew. Math. **471**, 23–42 (1996)

[APS] M. Atiyah, V. Patodi, I. Singer, Spectral asymmetry and Riemannian Geometry. I. Math. Proc. Camb. Philos. Soc. **77**, 43–69 (1975)

[B] C. Bär, Real killing spinors and holonomy. Commun. Math. Phys. **154**, 509–521 (1993)

[BB] C. Bär, W. Ballmann, Boundary value problems for elliptic differential operators of first order, in *Survey in Differential Geometry*, vol. 17, ed. by H.-D. Cao, S.-T. Yau (International Press, Somerville, 2012), pp. 1–78

[BBC] W. Ballmann, J. Brüning, G. Carron, Regularity and index theory for Dirac-Schrödinger systems with Lipschitz coefficients. J. Math. Pures Appl. **89**, 429–476 (2008)

[BW] B. Booss–Bavnbek, K. Wojciechowski, *Elliptic Boundary Problems for Dirac Operators* (Birkhäuser, Boston, 1993)

[Fr] D. Freed, Two index theorems in odd dimensions. Commun. Anal. Geom. **6**, 317–329 (1998)

[Fri] T. Friedrich, Der erste Eigenwert des Dirac-Operators einer kompakten, Riemannschen Mannigfaltigkeit nichtnegativer Skalarkrümmung. Math. Nachr. **97**, 117–146 (1980)

[Gi] P. Gilkey, *Invariance Theory, the Heat Equation, and the Atiyah-Singer Index Theorem* (Publish or Perish, Wilmington, 1984)

[GL] M. Gromov, H.B. Lawson, Positive scalar curvature and the Dirac operator on complete Riemannian manifolds. Inst. Hautes Études Sci. Publ. Math. **58**(1983), 83–196 (1984)

[GN] N. Große, R. Nakad, Boundary value problems for noncompact boundaries of spin^c manifolds and spectral estimates. arxiv:1207.4568[v2] Proc. Lond. Math. Soc. (3)109 (2014), **4**, 946–974

[Hi] N. Higson, A note on the cobordism invariance of the index. Topology **30**, 439–443 (1991)

[HMR] O. Hijazi, S. Montiel, A. Roldán, Eigenvalue boundary problems for the Dirac operator. Commun. Math. Phys. **231**, 375–390 (2002)

[Ka] T. Kato, *Perturbation Theory for Linear Operators*. 2nd edn. (Springer, Berlin, 1986)

[LM] H.B. Lawson, M.-L. Michelsohn, *Spin Geometry* (Princeton University Press, Princeton, 1989)

[Pa] R.S. Palais, *Seminar on the Atiyah-Singer Index Theorem*. With contributions by M. F. Atiyah, A. Borel, E. E. Floyd, R. T. Seeley, W. Shih, and R. Solovay (Princeton University Press, Princeton, 1965)

[Ro] J. Roe, Partitioning noncompact manifolds and the dual Toeplitz problem, in *Operator Algebras and Applications*, vol. 1. London Mathematical Society Lecture Note Series, vol. 135 (Cambridge University Press, Cambridge, 1988), pp. 187–228

[Se] R.T. Seeley, Complex powers of an elliptic operator. Proc. Symp. Pure Math. **10**, 288–307 (1967)

[Ta] M.E. Taylor, *Partial Differential Equations. I. Basic Theory* (Springer, New York, 1996)

Symplectic and Hyperkähler Implosion

Andrew Dancer, Brent Doran, Frances Kirwan, and Andrew Swann

Abstract We review the quiver descriptions of symplectic and hyperkähler implosion in the case of SU(n) actions. We give quiver descriptions of symplectic implosion for other classical groups, and discuss some of the issues involved in obtaining a similar description for hyperkähler implosion.

2000 *Mathematics Subject Classification.* 53C26, 53D20, 14L24.

1 Introduction

Symplectic implosion is an abelianisation construction in symplectic geometry invented by Guillemin et al. [GJS]. Given a symplectic manifold M with a Hamiltonian action of a compact group K, its imploded cross-section M_{impl} is a

A. Dancer
Jesus College, Oxford OX1 3DW, UK
e-mail: dancer@maths.ox.ac.uk

B. Doran
Department of Mathematics, ETH Zürich, 8092 Zürich, Switzerland
e-mail: brent.doran@math.ethz.ch

F. Kirwan (✉)
Balliol College, Oxford OX1 3BJ, UK
e-mail: kirwan@maths.ox.ac.uk

A. Swann
Department of Mathematics, Aarhus University, Ny Munkegade 118, Bldg 1530, DK-8000 Aarhus C, Denmark

CP^3-Origins, Centre of Excellence for Cosmology and Particle Physics Phenomenology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark
e-mail: swann@math.au.dk

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_4

symplectic stratified space with a Hamiltonian action of a maximal torus T of K, such that the symplectic reductions of M by K agree with the symplectic reductions of the implosion by T.

There is a universal example of symplectic implosion, obtained by taking M to be the cotangent bundle T^*K. The imploded space $(T^*K)_{\mathrm{impl}}$ carries a Hamiltonian torus action for which the symplectic reductions are the coadjoint orbits of K. It also carries a Hamiltonian action of K which commutes with the T action, and the implosion M_{impl} of any symplectic manifold M with a Hamiltonian action of K can be constructed as the symplectic reduction at 0 of the product $M \times (T^*K)_{\mathrm{impl}}$ by the diagonal action of K.

The universal symplectic implosion $(T^*K)_{\mathrm{impl}}$ can also be described in a more algebraic way, as the geometric invariant theory quotient $K_{\mathbb{C}}//N$ of the complexification $K_{\mathbb{C}}$ of K by a maximal unipotent subgroup N. This is the affine variety $\mathrm{Spec}(\mathcal{O}(K_{\mathbb{C}})^N)$ associated with the algebra of N-invariant regular functions on $K_{\mathbb{C}}$, and may also be described as the canonical affine completion of the orbit space $K_{\mathbb{C}}/N$ which is a dense open subset of $K_{\mathbb{C}}//N$.

Many constructions in symplectic geometry involving the geometry of moment maps have analogues in hyperkähler geometry. We recall here that a hyperkähler structure is given by a Riemannian metric g and a triple of complex structures satisfying the quaternionic relations. In fact we then acquire a whole two-sphere's worth of complex structures, parametrised by the unit sphere in the imaginary quaternions. The metric is required to be Kähler with respect to each of the complex structures. In this way a hyperkähler structure defines a two-sphere of symplectic structures.

Just as the cotangent bundle T^*K of a compact Lie group carries a natural symplectic structure, so, by work of Kronheimer, the cotangent bundle $T^*K_{\mathbb{C}}$ of the complexified group carries a hyperkähler structure [Kro1]. Moreover, in a series of papers Kronheimer, Biquard and Kovalev showed that the coadjoint orbits of $K_{\mathbb{C}}$ admit hyperkähler structures [Kro2, Kro3, Biq, Kov]. These orbits are not, however, closed in $\mathfrak{k}_{\mathbb{C}}^*$ (and the hyperkähler metrics are not complete) except in the case of semisimple orbits.

In [DKS1] and subsequent papers [DKS2, DKS3] we developed a notion of a universal hyperkähler implosion $(T^*K_{\mathbb{C}})_{\mathrm{hkimpl}}$ for $SU(n)$ actions. The hyperkähler implosion of a general hyperkähler manifold M with a Hamiltonian action of $K = SU(n)$ can then be defined as the hyperkähler quotient of $M \times (T^*K_{\mathbb{C}})_{\mathrm{hkimpl}}$ by the diagonal action of K. As in the symplectic case the universal hyperkähler implosion carries an action of $K \times T$ where $K = SU(n)$ and T is its standard maximal torus. As coadjoint orbits for the complex group are no longer closed in general, and are not uniquely determined by eigenvalues, the hyperkähler quotients of $(T^*K_{\mathbb{C}})_{\mathrm{hkimpl}}$ by the torus action need not be single orbits. Instead, they are the Kostant varieties, that is, the varieties in $\mathfrak{sl}(n, \mathbb{C})^*$ obtained by fixing the values of the invariant polynomials for this Lie algebra. These varieties are unions of coadjoint orbits and are closures in $\mathfrak{sl}(n, \mathbb{C})^*$ of the regular coadjoint orbits of $K_{\mathbb{C}} = SL(n, \mathbb{C})$. We refer to [CG, Kos] for more background on the Kostant varieties.

Again by analogy with the symplectic case, we can describe the hyperkähler implosion in terms of geometric invariant theory (GIT) quotients by non-reductive group actions. Explicitly, the implosion is $(SL(n,\mathbb{C}) \times \mathfrak{n}^0) /\!\!/ N$ where N is a maximal unipotent subgroup of $K_{\mathbb{C}} = SL(n,\mathbb{C})$ and $\mathfrak{n}^0$ is the annihilator in $\mathfrak{sl}(n,\mathbb{C})^*$ of its Lie algebra $\mathfrak{n}$. Thus the universal hyperkähler implosion for $K = SU(n)$ can be identified with the complex-symplectic quotient $(SL(n,\mathbb{C}) \times \mathfrak{n}^\circ) /\!\!/ N$ of $T^* SL(n,\mathbb{C})$ by N in the GIT sense, just as the symplectic implosion is the GIT quotient of $K_{\mathbb{C}}$ by N.

In the case of $K = SU(n)$ dealt with in [DKS1], it is possible to describe the hyperkähler implosion via a purely finite-dimensional construction using quiver diagrams. This construction was motivated by a quiver description of the symplectic implosion for $SU(n)$ we described in §4 of [DKS1].

In this article we shall extend these results concerning symplectic implosion to other classical groups, that is, the special orthogonal and symplectic groups. Our approach will be inspired by the description by Lian and Yau [LY] of coadjoint orbits for compact classical groups using quivers. This suggests ways to extend the quiver construction of the universal hyperkähler implosion from the case of $SU(n)$ to general classical groups, and we discuss some of the issues involved here.

Acknowledgements We thank Kevin McGerty and Tom Nevins for bringing the paper [GR] to our attention. The second author is partially supported by Swiss National Science Foundation grant 200021 138071. The fourth author is partially supported by the Danish Council for Independent Research, Natural Sciences.

2 Symplectic Quivers

We begin by trying to construct a quiver model for the universal symplectic implosion in the case of the orthogonal and symplectic groups, as was done in [DKS1] for special unitary groups.

We consider diagrams of vector spaces and linear maps

$$0 = V_0 \stackrel{\alpha_0}{\to} V_1 \stackrel{\alpha_1}{\to} V_2 \stackrel{\alpha_2}{\to} \cdots \stackrel{\alpha_{r-2}}{\to} V_{r-1} \stackrel{\alpha_{r-1}}{\to} V_r = \mathbb{C}^n. \tag{1}$$

The dimension vector is defined to be $\mathbf{n} = (n_1, \ldots, n_{r-1}, n_r = n)$ where $n_i = \dim V_i$. We will say that the representation is *ordered* if $0 \leqslant n_1 \leqslant n_2 \leqslant \cdots \leqslant n_r = n$ and *strictly ordered* if $0 < n_1 < n_2 < \cdots < n_r = n$.

In [DKS1] we considered $V_r = \mathbb{C}^n$ as a representation of $SU(n)$, or its complexification $SL(n,\mathbb{C})$. In this setting we say the quiver is *full flag* if $r = n$ and $n_i = i$ for each i. We took the geometric invariant theory quotient of the space of full flag quivers by $SL := \prod_{i=1}^{r-1} SL(V_i)$ (or equivalently, the symplectic quotient by $\prod_{i=1}^{r-1} SU(V_i)$). The stability conditions imply that the quiver decomposes into a quiver with zero maps and a quiver with all maps injective. It was therefore sufficient to analyse the injective quivers up to equivalence. We found that the quotient could be stratified into 2^{n-1} strata, indexing the flag of dimensions of the injective quivers (after the quivers had been contracted to remove edges where the maps were isomorphisms). Equivalently, the strata were indexed by the ordered partitions of n. Each stratum could be identified with $SL(n,\mathbb{C})/[P,P]$ where P is the parabolic associated with the given flag. The upshot was that the full GIT quotient can be identified as an affine variety with the affine completion $SL(n,\mathbb{C})/\!\!/N = \operatorname{Spec}\mathcal{O}(SL(n,\mathbb{C}))^N$ of the open stratum $SL(n,\mathbb{C})/N$.

We now wish to view V_r as a representation of an orthogonal or symplectic group. This involves introducing the associated bilinear forms. Our approach will be motivated by the description due to Lian and Yau in [LY] of a quiver approach to generalised flag varieties for symplectic and orthogonal groups.

Note that for consistency with [DKS1] we have altered the notation of [LY] in some respects. In particular we use r rather than $r+1$ for the top index, and we use n_i rather than d_i for the dimensions.

In the orthogonal case, we let J denote the matrix with entries

$$J_{ij} = \delta_{n+1-i,j} \qquad (1 \leqslant i,j \leqslant n)$$

which are 1 on the antidiagonal and 0 elsewhere. We therefore have on $\mathbb{C}^n$ a symmetric bilinear form $B(v,w) = v^tJw$ where v^t denotes the transpose of v, which is preserved by

$$SO(n,\mathbb{C}) = \{\, g : g^tJg = J \,\}$$

We note that the condition for h to be in the Lie algebra $\mathfrak{so}(n,\mathbb{C})$ is $h^tJ + Jh = 0$, that is, that h is skew-symmetric about the ANTI-diagonal. In particular, h may have arbitrary elements in the top left $d \times d$ block as long as $d \leqslant \frac{n}{2}$.

Motivated by [LY] let us now consider ordered diagrams where

$$n_{r-1} \leqslant \frac{n}{2}$$

and we impose on α_{r-1} the condition

$$\alpha_{r-1}^t J \alpha_{r-1} = 0. \tag{2}$$

Equivalently, this is the condition that the image of α_{r-1} be an isotropic subspace of $\mathbb{C}^n$ with respect to J (which is the reason for the inequality above). The space of α_{r-1} satisfying this condition is of dimension $nn_{r-1} - \frac{1}{2}n_{r-1}(n_{r-1}+1)$.

We let $R(\mathbf{n})$ be the space of all such diagrams satisfying (2) with dimension vector $\mathbf{n}$.

Observe that the complexification $GL := \prod_{i=1}^{r-1} GL(V_i)$ of $\tilde{H} := \prod_{i=1}^{r-1} U(V_i)$ acts on $R(\mathbf{n})$ by

$$\begin{aligned} \alpha_i &\mapsto g_{i+1}\alpha_i g_i^{-1} \quad (i = 1,\ldots,r-2), \\ \alpha_{r-1} &\mapsto \alpha_{r-1} g_{r-1}^{-1}. \end{aligned}$$

There is also a commuting action of $SO(n, \mathbb{C})$ by left multiplication of α_{r-1}; note that the full group $GL(n, \mathbb{C})$ does not now act because it does not preserve (2).

As in [DKS1], we shall study the symplectic quotient of $R(\mathbf{n})$ by the action of

$$H := \prod_{i=1}^{r-1} SU(V_i)$$

or equivalently the GIT quotient of $R(\mathbf{n})$ by its complexification

$$H_{\mathbb{C}} = SL := \prod_{i=1}^{r-1} SL(V_i),$$

viewed as a subgroup of GL in the obvious way. This quotient will have residual actions of the $r-1$-dimensional compact torus $T^{r-1} = (S^1)^{r-1}$ and its complexification, as well as of $SO(n, \mathbb{C})$.

Let us observe that the dimension $nn_{r-1} - \frac{1}{2}n_{r-1}(n_{r-1}+1)$ of the set of α_{r-1} satisfying (2) equals the dimension of the coset space $SO(n, \mathbb{C})/\, SO(n - n_{r-1}, \mathbb{C})$. In fact, in the orthogonal case with n odd, we can show that this coset space equals the set of injective α_{r-1} satisfying (2).

For if n is odd $SO(n, \mathbb{C})$ acts transitively on the set of isotropic subspaces of $\mathbb{C}^n$ of fixed dimension, so α_{r-1} can be put into the form

$$\begin{pmatrix} A_{n_{r-1} \times n_{r-1}} \\ 0_{(n-n_{r-1}) \times n_{r-1}} \end{pmatrix}$$

via the $SO(n, \mathbb{C})$ action. As $n_{r-1} \leqslant \frac{n}{2}$, we can consider matrices in $SO(n, \mathbb{C})$ with an arbitrary invertible $n_{r-1} \times n_{r-1}$ block in the top left corner and a zero $(n-n_{r-1}) \times n_{r-1}$ block in the lower left. So in fact α_{r-1} can be put into the standard form (used also in the A_n case in [DKS1])

$$\begin{pmatrix} I_{n_{r-1} \times n_{r-1}} \\ 0_{(n-n_{r-1}) \times n_{r-1}} \end{pmatrix}.$$

The connected component of the stabiliser of this configuration for the $SO(n, \mathbb{C})$ action is $SO(n - n_{r-1}, \mathbb{C})$.

We now obtain a description of quivers in $R(\mathbf{n})$ with all α_i injective, modulo the action of SL. For, combining the above observation with the arguments of §4 of [DKS1], the action of $SL \times SO(n, \mathbb{C})$ can be used to put the maps in standard form

$$\alpha_i = \begin{pmatrix} I_{n_i \times n_i} \\ 0_{(n_{i+1}-n_i) \times n_i} \end{pmatrix}.$$

The remaining freedom is a commutator of a parabolic in $SO(n, \mathbb{C})$ where the first $r-1$ block sizes are $n_{i+1} - n_i$. The blocks in the Levi subgroup lie in $SL(n_{i+1} - n_i)$ which is why we get the commutator rather than the full parabolic. Hence the injective quivers with fixed dimension vector $\mathbf{n}$ modulo the action of SL are parametrised by $SO(n, \mathbb{C})/[P, P]$, where P is the parabolic associated with the dimension vector.

Note that the blocks corresponding to the upper left square of size $n_{r-1} = \sum_{i=0}^{r-2} n_{i+1} - n_i$ will determine the blocks in the lower right square of size n_{r-1}, at least on Lie algebra level, by the property of being in the orthogonal group.

Remark 2.1. By intersecting the parabolic (rather than its commutator) with the compact group $SO(n)$ we get $\prod_{i=0}^{r-2} U(n_{i+1}-n_i) \times SO(n-n_{r-1})$, which is the isotropy group for the associated compact flag variety. Putting $p_{i+1} = n_{i+1} - n_i$, and $\ell = n - n_{r-1}$, we get $\sum_{i=1}^{r-1} p_i = n - l$, in accordance with the results of [Bes, p. 233, Sect. 8H].

For a model for the non-reductive GIT quotient $K_{\mathbb{C}}//N$ in the B_k case, that is when $n = 2k+1$ and $K = SO(2k+1)$, we can take $\mathbf{n}$ equal to $(1, 2, 3, \ldots, k, 2k+1)$ which will be the full flag condition in this context. We now consider the GIT quotient $R(\mathbf{n})/\!\!/ SL$. Note that at this stage the maps α_i are not assumed to be injective.

As the SL action is the same as in the A_n case, the stability analysis proceeds as in [DKS1]. We find that for polystable configurations we may decompose each vector space $\mathbb{C}^i$ as

$$\mathbb{C}^i = \ker \alpha_i \oplus \mathbb{C}^{m_i}, \tag{3}$$

where $\mathbb{C}^{m_i} = \operatorname{im} \alpha_{i-1}$ if $m_i \neq 0$. So the quiver decomposes into a zero quiver and an injective quiver. After contracting legs of the quiver which are isomorphisms, as in [DKS1], we obtain a strictly ordered injective quiver of the form considered above. We stratify the quotient $R(\mathbf{n})/\!\!/ SL$ by the flag of dimensions of the injective quiver, as in the $SL(n, \mathbb{C})$ case.

We have thus identified the strata of the GIT quotient of the space of full flag quivers with the strata $SO(n, \mathbb{C})/[P, P]$ of the universal symplectic implosion, or equivalently the non-reductive GIT quotient $K_{\mathbb{C}}//N$ (where N is a maximal unipotent subgroup). As the complement of the open stratum $SO(n, \mathbb{C})/N$ is of complex codimension strictly greater than one, we see that the implosion $K_{\mathbb{C}}//N$ and the GIT quotient of the space of full flag quivers are affine varieties with the same coordinate ring $\mathcal{O}(SO(n, \mathbb{C}))^N$, and so are isomorphic.

Guillemin et al. [GJS] showed that the non-reductive GIT quotient $K_{\mathbb{C}}//N$ has a $K \times T$-invariant Kähler structure such that it can be identified symplectically with the universal symplectic implosion for K. In order to see this Kähler structure on $R(\mathbf{n})// SL$ we can put an $\tilde{H} \times K$-invariant flat Kähler structure on $R(\mathbf{n})$ and identify the GIT quotient $R(\mathbf{n})// SL$ with the symplectic quotient $R(\mathbf{n})/\!\!/ H$. To achieve $\tilde{H} \times K$-invariance we use the standard flat Kähler structure on $(\mathbb{C}^{j-1})^* \otimes \mathbb{C}^j$ for $j \leqslant r-1$ but the flat Kähler structure defined by J on

$$(\mathbb{C}^{r-1})^* \otimes \mathbb{C}^n \cong (\mathbb{C}^n)^{r-1}.$$

Recall that a polystable quiver decomposes into the sum of a zero quiver and an injective quiver, and determines for us a partial flag in $\mathbb{C}^n$

$$W_1 \subseteq W_2 \subseteq \cdots \subseteq W_{r-1} \subseteq \mathbb{C}^n,$$

where $W_j = \operatorname{im} \alpha_{r-1} \circ \alpha_{r-2} \circ \cdots \circ \alpha_j$, whose dimension vector $(w_1 = \dim W_1, \dots, w_{r-1} = \dim W_{r-1})$ is determined by the injective summand. The condition that $\alpha_{r-1}^t J \alpha_{r-1} = 0$ ensures that W_{r-1} is an isotropic subspace of $\mathbb{C}^n$, so we can use the action of the compact group $K = SO(n)$ to put this flag into standard form with W_j spanned by the first w_j vectors in the standard basis for $\mathbb{C}^n$. Then we can use the action of SL to put the polystable quiver into the form where

$$\alpha_j = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \nu_1^j & 0 & \dots & 0 \\ 0 & \nu_2^j & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \nu_j^j \end{pmatrix} \tag{4}$$

for $j < r-1$ and

$$\alpha_{r-1} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \\ \nu_1^{r-1} & 0 & \dots & 0 \\ 0 & \nu_2^{r-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \nu_{r-1}^{r-1} \end{pmatrix}$$

with $\nu_i^j \in \mathbb{C}$. Let $R_T(\mathbf{n})$ denote the subspace of $R(\mathbf{n})$ consisting of quivers of this form.

Note that the moment map for the action of the unitary group $U(V_j)$ on $R(\mathbf{n})$ takes a quiver (1) to

$$\bar{\alpha}_j^t \alpha_j - \alpha_{j-1} \bar{\alpha}_{j-1}^t$$

for $1 \leqslant j \leqslant r-1$, so the moment map for the action of the product $\tilde{H} = \prod_{j=1}^{r-1} U(V_j)$ takes $R_T(\mathbf{n})$ into the Lie algebra of the product of the standard (diagonal) maximal tori T_{V_j} of the unitary groups $U(V_j)$. Thus there is a natural map of symplectic quotients

$$\theta_T \colon R_T(\mathbf{n}) /\!/ H_T \to R(\mathbf{n}) /\!/ H$$

where $H_T = \prod_{j=1}^{r-1}(T_{V_j} \cap SL(V_j))$ is a maximal torus of H. Moreover

$$R(\mathbf{n})/\!\!/H = K\,\theta_T(R_T(\mathbf{n})/\!\!/H_T)$$

where $K = SO(n)$ and $R_T(\mathbf{n})/\!\!/H_T$ is a toric variety.

The moment map for the action of the torus $T^{r-1} = (S^1)^{r-1}$ on $R(\mathbf{n})/\!\!/H$ takes a point represented by a quiver of the form (4) satisfying the moment map equations

$$\begin{pmatrix} |v_1^j|^2 & 0 & \dots & 0 \\ 0 & |v_2^j|^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |v_j^j|^2 \end{pmatrix} = \lambda_j^{\mathbb{R}} I + \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & |v_1^{j-1}|^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |v_{j-1}^{j-1}|^2 \end{pmatrix}$$

for some $\lambda_1^{\mathbb{R}}, \dots, \lambda_{r-1}^{\mathbb{R}} \in \mathbb{R}$, or equivalently

$$|v_i^j|^2 = \lambda_j^{\mathbb{R}} + \lambda_{j-1}^{\mathbb{R}} + \dots + \lambda_{j-i+1}^{\mathbb{R}} \qquad \text{if } 1 \leqslant i \leqslant j < n, \tag{5}$$

to $(\lambda_1^{\mathbb{R}}, \dots, \lambda_{r-1}^{\mathbb{R}})$ in the Lie algebra of T^{r-1}, while the moment map for the action of K takes this point to

$$\begin{pmatrix} -|v_{r-1}^{r-1}|^2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & -|v_1^{r-1}|^2 & 0 & \dots & 0 \\ 0 & \dots & 0 & |v_1^{r-1}|^2 & \dots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & |v_{r-1}^{r-1}|^2 \end{pmatrix},$$

up to constant scalar factors depending on conventions. The image of the toric variety $R_T(\mathbf{n})/\!\!/H_T$ under this moment map is the positive Weyl chamber $\mathfrak{t}_+$ of $K = SO(n)$, and we obtain a symplectic identification of $R(\mathbf{n})/\!\!/H$ with the universal symplectic implosion

$$(T^*K)_{\text{impl}} = (K \times \mathfrak{t}_+)/\sim$$

of $K = SO(n)$, where $(k, \xi) \sim (k', \xi')$ if and only if $\xi = \xi'$, with stabiliser K_ξ under the coadjoint action of K, and $k = k'\tilde{k}$ for some $\tilde{k} \in [K_\xi, K_\xi]$.

We may argue in a very similar way for the symplectic group $Sp(2k, \mathbb{C})$, the complexification of $Sp(2k)$. Following [LY] we replace the symmetric bilinear form by the skew form on $\mathbb{C}^n = \mathbb{C}^{2k}$ defined by the matrix

$$J_2 = \begin{pmatrix} 0 & J \\ -J & 0 \end{pmatrix}$$

Once again we find that $Sp(2k, \mathbb{C})$ acts transitively on the set of α_{r-1} satisfying the condition

$$\alpha_{r-1}^t J_2 \alpha_{r-1} = 0$$

and the above arguments go through mutatis mutandis.

For $SO(n, \mathbb{C})$ with n even, the isotropic subspace $\operatorname{im}\alpha_r$ may be self-dual or anti-self-dual if $n_{r-1} = \frac{n}{2}$. We take the component of the locus defined by (2) corresponding to the image being self-dual, and now we get the desired transitivity.

Theorem 2.2. *Let K be a compact classical group and let us consider full flag quivers for $K_{\mathbb{C}}$ as above. That is, we take $\boldsymbol{n} = (n_1, \ldots, n_r)$ to be $(1, 2, \ldots, n)$ for $K = SU(n)$, $(1, 2, \ldots, k, 2k+1)$ for $SO(2k+1)$, and $(1, 2, \ldots, k, 2k)$ for $SO(2k)$ or $Sp(k)$. Also in the orthogonal and symplectic cases we impose the appropriate isotropy condition on the top map in the quiver, and take the appropriate component of the space of isotropic subspaces in the even orthogonal case, to obtain a space $R(\mathbf{n})$ of full flag quivers.*

Then the symplectic quotient of $R(\mathbf{n})$ by $H(\mathbf{n}) = \prod_{i=2}^{r-1} SU(n_i, \mathbb{C})$ can be identified naturally with the universal symplectic implosion for K, or equivalently with the non-reductive GIT quotient $K_{\mathbb{C}}//N$. The stratification by quiver diagrams as above corresponds to the stratification of the universal symplectic implosion as the disjoint union over the standard parabolic subgroups P of $K_{\mathbb{C}}$ of the varieties $K_{\mathbb{C}}/[P, P]$.

Example 2.3. The lowest rank case of the above construction is when the group is $SO(3)$. The quiver is now just

$$0 \stackrel{\alpha_0}{\to} \mathbb{C} \stackrel{\alpha_1}{\to} \mathbb{C}^3$$

where $\alpha_0 = 0$. As $H = SU(1)$ here there is no quotienting to perform. The matrix J is $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ and, putting $\alpha_1 = (x, y, z)$, the isotropy condition $\alpha_1^t J \alpha_1 = 0$ becomes

$$y^2 + xz = 0.$$

This affine surface is a well-known description of the Kleinian singularity $\mathbb{C}^2/\mathbb{Z}_2$. This is a valid description of the symplectic implosion for $SO(3)$, since the implosion for the double cover $SU(2)$ is just $\mathbb{C}^2$.

Remark 2.4. We mention here an alternative description of symplectic implosions using the concept of Cox rings [Cox, HK, LV]. If X is an algebraic variety and $L_1, \ldots, L_n$ are generators for $\operatorname{Pic}(X)$, then we form the Cox ring

$$\mathrm{Cox}(X, L) = \bigoplus_{(m_1,\ldots,m_n)\in\mathbb{Z}^n} H^0(X, m_1L_1 + \cdots + m_nL_n) \tag{6}$$

Hu and Keel [HK] introduced the class of *Mori dream spaces*—the varieties X whose Cox ring is finitely generated. These include toric varieties, which are characterised by $\mathrm{Cox}(X)$ being a polynomial ring. It was proved in [HK] that, as the name suggests, Mori dream spaces are well behaved from the point of view of the Minimal Model Programme. After a finite sequence of flips and divisorial contractions we arrive at a space birational to X which either is a Mori fibre space or has nef canonical divisor. Mori dream spaces may be realised as GIT quotients by tori of the affine varieties associated with their Cox rings. The above sequence of flips and contractions can be expressed in terms of explicit variation of GIT wall-crossings, and indeed the Mori chambers admit a natural identification with variation of GIT chambers. Since torus variation of GIT is well-understood, in principle the Mori theory of a Mori dream space is also well-understood, at least given an explicit enough presentation of Cox(X).

If K is a compact Lie group and P is a parabolic subgroup of $K_{\mathbb{C}}$, then the Cox ring of $K_{\mathbb{C}}/P$ is the coordinate ring of the quasi-affine variety $K_{\mathbb{C}}/[P, P]$, which is finitely generated. In particular $K_{\mathbb{C}}/P$ is a Mori dream space.

Taking P to be a Borel subgroup B, we find that the Cox ring of $K_{\mathbb{C}}/B$ is the finitely generated ring $\mathcal{O}(K_{\mathbb{C}}/N) = \mathcal{O}(K_{\mathbb{C}})^N$ whose associated affine variety $\mathrm{Spec}(\mathcal{O}(K_{\mathbb{C}})^N)$ is the universal symplectic implosion $K_{\mathbb{C}}//N$.

3 Hyperkähler Quiver Diagrams

For $K = SU(n)$ actions we developed in [DKS1] a finite-dimensional approach to constructing the universal hyperkähler implosion for K via quiver diagrams. In that case the symplectic quivers formed a linear space and we just took the cotangent bundle, which amounted to putting in maps $\beta_i\colon V_{i+1} \to V_i$ in addition to the $\alpha_i\colon V_i \to V_{i+1}$. Writing $V_i = \mathbb{C}^{n_i}$, we thus worked with the flat hyperkähler space

$$M = M(\mathbf{n}) = \bigoplus_{i=1}^{r-1} \mathbb{H}^{n_i n_{i+1}} = \bigoplus_{i=1}^{r-1} \mathrm{Hom}(\mathbb{C}^{n_i}, \mathbb{C}^{n_{i+1}}) \oplus \mathrm{Hom}(\mathbb{C}^{n_{i+1}}, \mathbb{C}^{n_i}) \tag{7}$$

with the hyperkähler action of $U(n_1) \times \cdots \times U(n_r)$

$$\alpha_i \mapsto g_{i+1}\alpha_i g_i^{-1}, \quad \beta_i \mapsto g_i \beta_i g_{i+1}^{-1} \qquad (i = 1, \ldots r-1),$$

with $g_i \in U(n_i)$ for $i = 1, \ldots, r$. Right quaternion multiplication was given by

$$(\alpha_i, \beta_i)\mathbf{j} = (-\beta_i^*, \alpha_i^*). \tag{8}$$

If each β_i is zero we recovered a symplectic quiver diagram.

We considered the hyperkähler quotient of $M(\mathbf{n})$ with respect to the group $H = \prod_{i=1}^{r-1} SU(n_i)$, obtaining a stratified hyperkähler space $Q = M/\!/\!/H$, which has a residual action of the torus $T^{r-1} = \tilde{H}/H$ where $\tilde{H} = \prod_{i=1}^{r-1} U(n_i)$, as well as a commuting action of $SU(n_r) = SU(n)$. When $\mathbf{n} = (1, 2, \ldots, n)$ we can identify this torus with the standard maximal torus T of $SU(n)$ using the simple roots of T.

The *universal hyperkähler implosion for* $SU(n)$ is defined to be the hyperkähler quotient $Q = M/\!/\!/H$, where M, H are as above with $n_j = j$, for $j = 1, \ldots, n$, (i.e. the case of a full flag quiver).

From the complex-symplectic viewpoint, Q is the GIT quotient, by the complexification

$$H_{\mathbb{C}} = \prod_{i=1}^{r-1} SL(n_i, \mathbb{C})$$

of H, of the zero locus of the complex moment map $\mu_{\mathbb{C}}$ for the H action.

The components of this complex moment map $\mu_{\mathbb{C}}$ are given by the tracefree parts of $\alpha_{i-1}\beta_{i-1} - \beta_i\alpha_i$. The complex moment map equation $\mu_{\mathbb{C}} = 0$ can thus be expressed as saying

$$\beta_i\alpha_i - \alpha_{i-1}\beta_{i-1} = \lambda_i^{\mathbb{C}} I \qquad (i = 1, \ldots, r-1), \tag{9}$$

for some complex scalars $\lambda_1^{\mathbb{C}}, \ldots, \lambda_{r-1}^{\mathbb{C}}$, while the real moment map equation is given by

$$\beta_{i-1}^*\beta_{i-1} - \alpha_{i-1}\alpha_{i-1}^* - \beta_i\beta_i^* + \alpha_i^*\alpha_i = \lambda_i^{\mathbb{R}} I \qquad (i = 1, \ldots, r-1), \tag{10}$$

for some real scalars $\lambda_1^{\mathbb{R}}, \ldots, \lambda_{r-1}^{\mathbb{R}}$.

The action of $H_{\mathbb{C}}$ is given by

$$\alpha_i \mapsto g_{i+1}\alpha_i g_i^{-1}, \quad \beta_i \mapsto g_i\beta_i g_{i+1}^{-1} \qquad (i = 1, \ldots r-2),$$
$$\alpha_{r-1} \mapsto \alpha_{r-1} g_{r-1}^{-1}, \quad \beta_{r-1} \mapsto g_{r-1}\beta_{r-1},$$

where $g_i \in SL(n_i, \mathbb{C})$. The residual action of $SL(n, \mathbb{C}) = SL(n_r, \mathbb{C})$ on the quotient Q is given by

$$\alpha_{r-1} \mapsto g_r\alpha_{r-1}, \quad \beta_{r-1} \mapsto \beta_{r-1} g_r^{-1}.$$

There is also a residual action of $\tilde{H}_{\mathbb{C}}/H_{\mathbb{C}}$ which we can identify, in the full flag case, with the maximal torus $T_{\mathbb{C}}$ of $K_{\mathbb{C}}$. The complex numbers λ_i combine to give the complex-symplectic moment map for this complex torus action. We remark that reduction of Q by the maximal torus at level 0 recovers the construction of the nilpotent variety [KS, KP1]. While there is a similar quiver description of the nilpotent variety for the classical algebras $\mathfrak{so}(n, \mathbb{C})$ and $\mathfrak{sp}(n, \mathbb{C})$ [KS, KP2], the construction of an implosion does not directly generalise, partly because the corresponding groups $\tilde{H}$ do not have sufficiently large centres.

Note that as Q is a hyperkähler reduction by H at level 0, it also inherits an $SU(2)$ action that rotates the two-sphere of complex structures (see [DKS1] for details).

Given a quiver $(\alpha, \beta) \in M(\mathbf{n})$, the composition

$$X = \alpha_{r-1}\beta_{r-1} \in \mathrm{Hom}(\mathbb{C}^n, \mathbb{C}^n).$$

is invariant under the action of $\tilde{H}_{\mathbb{C}}$ and transforms by conjugation under the residual $SL(n, \mathbb{C})$ action. The map $Q \to \mathfrak{sl}(n, \mathbb{C})$ given by sending (α, β) to the tracefree part of X is therefore $T_{\mathbb{C}}$-invariant and $SL(n, \mathbb{C})$-equivariant.

In [DKS1] and [DKS2] we introduced stratifications of the implosion Q, one reflecting its hyperkähler structure and one reflecting the group structure of $SU(n)$. We recall in particular that the open subset of Q consisting of quivers with all β surjective may be identified with $SL(n, \mathbb{C}) \times_N \mathfrak{n}^0 \cong SL(n, \mathbb{C}) \times_N \mathfrak{b}$.

The open stratum Q^{hks} in the hyperkähler stratification of Q consists of the quivers which are hyperkähler stable; that is, for a generic choice of complex structure all the maps α_i are injective and all the maps β_i are surjective. In this situation the kernels of the compositions

$$\beta_j \circ \beta_{j+1} \circ \cdots \circ \beta_{n-1}$$

for $1 \leqslant j \leqslant n$ form a full flag in $\mathbb{C}^n$; we can use the action of $K = SU(n)$ (which preserves the hyperkähler structure) to put this flag into standard position. Next we can use the action of $SL = H_{\mathbb{C}}$ to put the maps β_j into the form

$$\beta_j = \begin{pmatrix} 0 & \mu_1^j & 0 & \dots & 0 & 0 \\ 0 & 0 & \mu_2^j & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \mu_{j-1}^j & 0 \\ 0 & 0 & 0 & \dots & 0 & \mu_j^j \end{pmatrix} \tag{11}$$

for some $\mu_i^j \in \mathbb{C} \setminus \{0\}$. Then it follows from the complex moment map equations (9) that the maps α_j have the form

$$\alpha_j = \begin{pmatrix} * & * & \dots & * & * \\ \nu_1^j & * & \dots & * & * \\ 0 & \nu_2^j & \dots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \nu_j^{j-1} & * \\ 0 & 0 & \dots & 0 & \nu_j^j \end{pmatrix} \tag{12}$$

and that the same Eqs. (9) are satisfied if each α_j is replaced with

$$\alpha_j^t = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \nu_1^j & 0 & \dots & 0 & 0 \\ 0 & \nu_2^j & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \nu_j^{j-1} & 0 \\ 0 & 0 & \dots & 0 & \nu_j^j \end{pmatrix} \tag{13}$$

For a fixed choice of complex structures let us denote by $\mathfrak{b}_+^{(\circ)}$ the subset of Q represented by all quivers of the form (12) and (11) satisfying the hyperkähler moment map equations with μ_i^j and ν_i^j nonzero complex numbers. Its K-sweep $K\,\mathfrak{b}_+^{(\circ)}$ in Q is then isomorphic to

$$K \times_T \mathfrak{b}_+^{(\circ)} \cong K_{\mathbb{C}} \times_B \mathfrak{b}_+^{(\circ)}$$

and consists of all quivers (1) in Q such that for each j the map α_j is injective and the map β_j is surjective and $\mathbb{C}^n$ is the direct sum of

$$\ker(\beta_j \circ \beta_{j+1} \circ \cdots \circ \beta_{n-1})$$

and $\mathrm{im}(\alpha_{n-1} \circ \cdots \circ \alpha_j)$. It follows that $K\,\mathfrak{b}_+^{(\circ)}$ is open in Q, and its sweep $SU(2)K\,\mathfrak{b}_+^{(\circ)}$ under the action of $SU(2)$ which rotates the complex structures (and commutes with the action of K) is the open stratum Q^{hks} of Q.

Associating to a quiver (1) in $\mathfrak{b}_+^{(\circ)}$ with the maps α_j and β_j in the form (12) and (11) the quiver in which α_j is replaced with α_j^T given by (13) defines a map ψ from $\mathfrak{b}_+^{(\circ)}$ to the hypertoric variety Q_T defined in [DKS2]. This hypertoric variety is the hyperkähler quotient of the space M_T of all quivers of the form (13) and (11) by the action of the maximal torus H_T of H with the induced action of $\tilde{H}_T/H_T = (S^1)^{n-1}$ which is identified with T via the basis of $\mathfrak{t}^* \cong \mathfrak{t}$ corresponding to the simple roots. The image of the map ψ is the open subset $Q_T^{(\circ)}$ of Q_T represented by all quivers of this form with μ_i^j and ν_i^j all nonzero.

The restriction to $\mathfrak{b}_+^{(\circ)}$ of the complex moment map for the action of K associates to a quiver of the form (12), (11) the upper triangular matrix

$$\alpha_{n-1}\beta_{n-1} - \mathrm{tr}(\alpha_{n-1}\beta_{n-1})\frac{I}{n}$$

and thus takes values in $\mathfrak{b} = \mathfrak{t}_{\mathbb{C}} \oplus \mathfrak{n}$. Combining the map ψ with the projection to $\mathfrak{n}$ of this complex moment map gives us isomorphisms

$$\mathfrak{b}_+^{(\circ)} \cong Q_T^{\circ} \times \mathfrak{n}$$

and

$$K\,\mathfrak{b}_+^{(\circ)} \cong K \times_T (Q_T^\circ \times \mathfrak{n}).$$

Under this identification the complex moment map for T is given by the (T-invariant) complex moment map $\phi\colon Q_T \to \mathfrak{t}_{\mathbb{C}}^*$ for the action of T on Q_T, and the complex moment map for K is given by

$$[k, \eta, \zeta] \mapsto \mathrm{Ad}^* k\,(\phi(\eta) + \zeta)$$

for $k \in K$, $\eta \in Q_T^\circ$ and $\zeta \in \mathfrak{n}$.

The hyperkähler moment map for T associates to a quiver satisfying the hyperkähler moment map equations (9) and (10) the element $(\lambda_1^{\mathbb{C}}, \lambda_1^{\mathbb{R}}, \ldots, \lambda_{n-1}^{\mathbb{C}}, \lambda_{n-1}^{\mathbb{R}})$ of $(\mathbb{C} \oplus \mathbb{R})^{n-1}$ identified with $\mathfrak{t}^* \otimes (\mathbb{C} \oplus \mathbb{R}) \cong \mathfrak{t}^* \otimes \mathbb{R}^3$ via the basis of simple roots. The image of its restriction to Q^{hks} is the open subset of $\mathfrak{t}^* \otimes \mathbb{R}^3$ defined by $(\lambda_j^{\mathbb{C}}, \lambda_j^{\mathbb{R}}) \neq (0,0)$ for $j = 1, \ldots, n-1$, while the image of $K\,\mathfrak{b}_+^{(\circ)}$ is the open subset $(\mathfrak{t}^* \otimes \mathbb{R}^3)^\circ$ defined by $\lambda_j^{\mathbb{C}} \neq 0$ for $j = 1, \ldots, n-1$. Using the same basis the hypertoric variety Q_T can be identified with $\mathbb{H}^{n-1}$ and $Q_T^{(\circ)}$ then corresponds to the open subset

$$\{\, (a_1 + jb_1, \ldots, a_{n-1} + jb_{n-1}) \in \mathbb{H}^{n-1} : a_\ell, b_\ell \in \mathbb{C} \setminus \{0\} \,\}.$$

Under this identification the hyperkähler moment map $\phi\colon Q_T^{(\circ)} \to \mathfrak{t}^* \otimes \mathbb{R}^3$ is given by

$$\begin{aligned} &\phi(a_1 + jb_1, \ldots, a_{n-1} + jb_{n-1}) \\ &\quad = (a_1 b_1, |a_1|^2 - |b_1|^2, \ldots, a_{n-1} b_{n-1}, |a_{n-1}|^2 - |b_{n-1}|^2); \end{aligned}$$

its fibres are single T-orbits in $Q_T^{(\circ)}$.

From the description of $K\,\mathfrak{b}_+^{(\circ)}$ above it follows that the hyperkähler moment map for T restricts to a locally trivial fibration

$$Q^{hks} \to SU(2)(\mathfrak{t}^* \otimes \mathbb{R}^3)^\circ$$

over the open subset $SU(2)(\mathfrak{t}^* \otimes \mathbb{R}^3)^\circ$ of $\mathfrak{t}^* \otimes \mathbb{R}^3$ with fibre $K \times \mathfrak{n}$.

Similarly the other strata in the hyperkähler stratification of Q are constructed from hyperkähler stable quivers of the form

$$0 \underset{\beta_0}{\overset{\alpha_0}{\rightleftarrows}} \mathbb{C}^{n_1} \underset{\beta_1}{\overset{\alpha_1}{\rightleftarrows}} \mathbb{C}^{n_2} \underset{\beta_2}{\overset{\alpha_2}{\rightleftarrows}} \cdots \underset{\beta_{r-2}}{\overset{\alpha_{r-2}}{\rightleftarrows}} \mathbb{C}^{n_{r-1}} \underset{\beta_{r-1}}{\overset{\alpha_{r-1}}{\rightleftarrows}} \mathbb{C}^{n_r} = \mathbb{C}^n. \tag{14}$$

Again for generic choices of complex structures for each j the map α_j is injective and the map β_j is surjective and $\mathbb{C}^n$ is the direct sum of

$$\ker(\beta_j \circ \beta_{j+1} \circ \cdots \circ \beta_{r-1})$$

and $\mathrm{im}(\alpha_{r-1} \circ \cdots \circ \alpha_j)$, and we can use the action of $K = SU(n)$ to put the flag in $\mathbb{C}^n$ defined by the subspaces $\ker(\beta_j \circ \beta_{j+1} \circ \cdots \circ \beta_{r-1})$ into standard position. Next we can use the action of $\prod_{j=1}^{r-1} SL(n_j)$ to put the maps β_j into block form of the same shape as (11) where now each μ_i^j is a nonzero scalar multiple of an identity matrix. Again it follows from the complex moment map equations (9) that the maps α_j have block form similar to (12) where each ν_i^j is a nonzero scalar multiple of an identity matrix, and that the same Eqs. (9) are satisfied if each α_j is replaced with α_j^T in block diagonal form as at (13). We find that the space Q_1^{hks} of hyperkähler quivers of the form (14) fibres over an open subset of $\mathfrak{t}_1^* \otimes \mathbb{R}^3$ (where $T_1 \cong (S^1)^{r-1}$) with fibre

$$K \times_{[K \cap P_1, K \cap P_1]} \mathfrak{p}_1^0$$

where P_1 is the standard parabolic in $K_{\mathbb{C}}$ corresponding to the flag defined by the subspaces $\ker(\beta_j \circ \beta_{j+1} \circ \cdots \circ \beta_{r-1})$, and $\mathfrak{p}_1^0$ is the annihilator in $\mathfrak{k}_{\mathbb{C}}^*$ of its Lie algebra $\mathfrak{p}_1$. Note that using the standard pairing on $\mathfrak{k}_{\mathbb{C}}$ and identifying $\mathfrak{b}$ with the annihilator $\mathfrak{n}^0$ of $\mathfrak{n}$ in $\mathfrak{k}_{\mathbb{C}}^*$, we have a projection from $\mathfrak{n} \cong \mathfrak{n}^*$ onto the annihilator of $\mathfrak{p}_1$ in $\mathfrak{n}^*$, and this annihilator can be identified with $\mathfrak{p}_1^0$ since $\mathfrak{n} + \mathfrak{p}_1 = \mathfrak{k}_{\mathbb{C}}$.

By [DKS1, Proposition 6.9] each stratum in Q can be identified with a hyperkähler modification

$$\hat{Q}_1^{hks} = (Q_1^{hks} \times (\mathbb{H} \setminus \{0\})^\ell) /\!/\!/ T^\ell$$

of Q_1^{hks} for some Q_1 as above, and the restriction to this stratum $\hat{Q}_1^{hks}$ of the hyperkähler moment map for T is a locally trivial fibration over an open subset of

$$\mathrm{Lie}(Z_K(K \cap P_1))^* \otimes \mathbb{R}^3$$

(where $Z_K(K \cap P_1) \subseteq T$ is the centre of $K \cap P_1$ in K) with fibre

$$K \times_{[K \cap P_1, K \cap P_1]} \mathfrak{p}_1^0 .$$

Using the surjection

$$K \times \mathfrak{n} \to K \times_{[K \cap P_1, K \cap P_1]} \mathfrak{p}_1^0 \qquad (15)$$

induced by the projection $\mathfrak{n} \cong \mathfrak{n}^* \to \mathfrak{p}_1^0$, we can lift this locally trivial fibration to one with fibre $K \times \mathfrak{n}$ which surjects onto the stratum $\hat{Q}_1^{hks}$.

In order to patch together these locally trivial fibrations for the different strata, we can blow up the hypertoric variety $Q_T \cong \mathbb{H}^{n-1}$, replacing it with $\tilde{Q}_T \cong \tilde{\mathbb{H}}^{n-1}$ where $\tilde{\mathbb{H}}^{n-1}$ is the blowup of $\mathbb{H} \cong \mathbb{C}^2$ at 0 using the complex structure on $\mathbb{H}$ given by right multiplication by i; this commutes with the hyperkähler complex structures given by left multiplication by i, j and k and also with the action of the S^1 component of the maximal torus $T \cong (S^1)^{n-1}$.

Note that the hyperkähler moment map for the action of S^1 on $\mathbb{H}$ induces an identification of the topological quotient $\mathbb{H}/S^1$ with $\mathbb{R}^3$; this pulls back to an identification of $\tilde{\mathbb{H}}/S^1$ with the manifold with boundary $\tilde{\mathbb{R}}^3 = (\mathbb{R}^3 \setminus \{0\}) \sqcup S^2$. Let $\tilde{Q}$ be the fibre product

$$\begin{array}{ccc} \tilde{Q} & \longrightarrow & Q \\ \downarrow & & \downarrow \\ \mathfrak{t}^* \otimes \tilde{R}^3 & \longrightarrow & \mathfrak{t}^* \otimes \mathbb{R}^3. \end{array}$$

The descriptions above of the hyperkähler strata of Q as the images of surjections from locally trivial fibrations over subsets of $\mathfrak{t}^* \otimes \mathbb{R}^3$ with fibre $K \times \mathfrak{n}$ patch together to give a locally trivial fibration

$$\hat{Q} \to \mathfrak{t}^* \otimes \tilde{\mathbb{R}}^3$$

with fibre $K \times \mathfrak{n}$ over the manifold with corners $\tilde{\mathbb{R}}^3$, and surjections $\hat{\chi}: \hat{Q} \to \tilde{Q}$ and $\chi: \tilde{Q} \to Q$ where $\hat{\chi}$ collapses fibres via the surjections (15) and χ is the pullback of the surjection $\mathfrak{t}^* \otimes \tilde{\mathbb{R}}^3 \to \mathfrak{t}^* \otimes \mathbb{R}^3$.

Remark 3.1. If at (11) we only allow ourselves to use the action of H, not $H_{\mathbb{C}}$, to put the maps β_j into standard form, then we are able to ensure that each β_j is of the form

$$\beta_j = \begin{pmatrix} 0 & \mu_1^j & * & \dots & * & * \\ 0 & 0 & \mu_2^j & \dots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \mu_{j-1}^j & * \\ 0 & 0 & 0 & \dots & 0 & \mu_j^j \end{pmatrix} \tag{16}$$

for some $\mu_i^j \in \mathbb{C} \setminus \{0\}$. It still follows from the complex moment map equations (9) that the maps α_j then have the form (12). Similarly an element of any stratum $\hat{Q}_1^{hks}$ as above can for generic choices of complex structures be put into block form (12) and (16) using the action of $K \times H$, where now μ_i^j and ν_i^j denote nonzero scalar multiples of identity matrices.

4 Properties of Hyperkähler Implosion

In this section we will list some of the main properties of the universal hyperkähler implosion $Q = (T^* K_{\mathbb{C}})_{\mathrm{hkimpl}}$ for $K = SU(n)$ which we expect to be true for more general compact groups K.

(1) Q is a stratified hyperkähler space of real dimension $2(\dim K + \dim T)$ where T is a maximal torus of K. It has an action of $K \times T$ which preserves the hyperkähler structure and has a hyperkähler moment map

$$\mu^{K\times T}: Q \to (\mathfrak{k}^* \oplus \mathfrak{t}^*) \otimes \mathbb{R}^3,$$

as well as a commuting action of $SU(2)$ which rotates the complex structures on Q (see [DKS1] for the case $K = SU(n)$).

(2) The hyperkähler reduction at 0 of Q by T can be identified for any choice of complex structure, via the complex moment map for the action of K, with the nilpotent cone $\mathcal{N}$ in $\mathfrak{k}_{\mathbb{C}}$. We can view this as the statement that the Springer resolution $SL(n,\mathbb{C}) \times_B \mathfrak{n} \to \mathcal{N}$ is an affinisation map. The reduction at a generic point of $\mathfrak{t}^* \otimes \mathbb{R}^3$ is a semisimple coadjoint orbit of $K_{\mathbb{C}}$, and in general the hyperkähler reduction of Q by T at any point of $\mathfrak{t}^* \otimes \mathbb{R}^3$ can be identified for any choice of complex structure, via the complex moment map for the action of K, with a Kostant variety in $\mathfrak{k}_{\mathbb{C}}^*$ (that is, the closure of a coadjoint orbit). We refer to [DKS1] for the case $K = SU(n)$.

(3) When K is semisimple, simply connected and connected (as for special unitary groups) its universal symplectic implosion embeds in the affine space

$$\bigoplus_{\varpi \in \Pi} V_{\varpi},$$

where $\{ V_{\varpi} : \varpi \in \Pi \}$ is the set of fundamental representations of K, as the closure of the $K_{\mathbb{C}}$-orbit of $v = \sum_{\varpi \in \Pi} v_{\varpi}$ for any choice of highest weight vector v_{ϖ} for the irreducible representation V_{ϖ}. When $K = SU(n)$ it was shown in [DKS3] that the universal hyperkähler implosion Q embeds in the space

$$H^0(\mathbb{P}^1, ((\mathfrak{k}_{\mathbb{C}}^* \oplus \mathfrak{t}_{\mathbb{C}}^*) \otimes \mathcal{O}(2)) \oplus \bigoplus_{\varpi} V_{\varpi} \otimes \mathcal{O}(j(\varpi)))$$

of holomorphic sections of the vector bundle

$$\mathcal{V} = ((\mathfrak{k}_{\mathbb{C}}^* \oplus \mathfrak{t}_{\mathbb{C}}^*) \otimes \mathcal{O}(2)) \oplus \bigoplus_{\varpi} V_{\varpi} \otimes \mathcal{O}(j(\varpi)) \tag{17}$$

over $\mathbb{P}^1$ for suitable positive integers $j(\varpi)$. Moreover this embedding induces a holomorphic and generically injective map from the twistor space $\mathcal{Z}Q$ of Q to the vector bundle $\mathcal{V}$ over $\mathbb{P}^1$, and the hyperkähler structure can be recovered from this embedding when $K = SU(n)$ [DKS3].

(4) Let N be a maximal unipotent subgroup of the complexification $K_{\mathbb{C}}$ of K. It was shown in [DKS1] that when $K = SU(n)$ the algebra of invariants $\mathcal{O}(K_{\mathbb{C}} \times \mathfrak{n}^0)^N$ is finitely generated and for any choice of complex structures Q is isomorphic to the affine variety

$$(K_{\mathbb{C}} \times \mathfrak{n}^0) /\!\!/ N = \operatorname{Spec} \mathcal{O}(K_{\mathbb{C}} \times \mathfrak{n}^0)^N$$

associated with this algebra of invariants. This variety may be viewed as the complex-symplectic quotient (in the sense of non-reductive GIT) of $T^*K_{\mathbb{C}} = K_{\mathbb{C}} \times \mathfrak{k}^*_{\mathbb{C}}$ by the action of N given by $(g, \zeta) \mapsto (gn^{-1}, Ad(n)\zeta)$. With respect to this identification the complex moment maps for the commuting K and T actions on Q are the morphisms from $(K_{\mathbb{C}} \times \mathfrak{n}^0) /\!\!/ N$ induced by the N-invariant morphisms from $K_{\mathbb{C}} \times \mathfrak{n}^0$ to $\mathfrak{k}^*_{\mathbb{C}}$ and $\mathfrak{t}^*_{\mathbb{C}}$ given by

$$(g, \zeta) \mapsto \mathrm{Ad}^*(g)\zeta$$

and

$$(g, \zeta) \mapsto \zeta_T$$

where $\zeta_T \in \mathfrak{t}^*_{\mathbb{C}}$ is the restriction of $\zeta \in \mathfrak{n}^0 \subseteq \mathfrak{k}^*_{\mathbb{C}}$ to $\mathfrak{t}_{\mathbb{C}}$.

It has been proved very recently by Ginzburg and Riche [GR, Lemma 3.6.2] that the algebra of regular functions on $T^*(G/N)$ is finitely generated for a general reductive G with maximal unipotent subgroup N. Taking $G = K_{\mathbb{C}}$ for any compact group K this cotangent bundle may be identified with $K_{\mathbb{C}} \times_N \mathfrak{n}^0$, and its algebra of regular functions is $\mathcal{O}(K_{\mathbb{C}} \times \mathfrak{n}^0)^N$. Hence the non-reductive GIT quotient $(K_{\mathbb{C}} \times \mathfrak{n}^0) /\!\!/ N = \operatorname{Spec} \mathcal{O}(K_{\mathbb{C}} \times \mathfrak{n}^0)^N$ is a well-defined affine variety in general, and is the canonical affine completion of the quasi-affine variety $T^*(K_{\mathbb{C}}/N)$ just as $K_{\mathbb{C}}//N$ is the canonical affine completion of the quasi-affine variety $K_{\mathbb{C}}/N$. It is enough to consider the case when K is semisimple, connected and simply connected. Then their proof provides a reasonably explicit set of generators involving the fundamental representations V_{ϖ} of K and these give an embedding of the affine variety $(K_{\mathbb{C}} \times \mathfrak{n}^0) /\!\!/ N = \operatorname{Spec} \mathcal{O}(K_{\mathbb{C}} \times \mathfrak{n}^0)^N$ as a closed subvariety of the space of sections $H^0(\mathbb{P}^1, \mathcal{V})$ of a vector bundle $\mathcal{V}$ over $\mathbb{P}^1$ as at (17) above. Note also that the GIT complex-symplectic quotient at level 0 of $(K_{\mathbb{C}} \times \mathfrak{n}^0) /\!\!/ N$ may be viewed as $(K_{\mathbb{C}} \times \mathfrak{n}) /\!\!/ B$ which is the nilpotent variety (see the remarks in (2) above). Similarly reductions at other levels will yield the Kostant varieties (cf. the discussion in §3 of [DKS1]).

Thus we expect that in general, as in the case when $K = SU(n)$, $(K_{\mathbb{C}} \times \mathfrak{n}^0) /\!\!/ N$ has a hyperkähler structure determined by this embedding and can be identified with the universal hyperkähler implosion for K.

Note that the scaling action of $\mathbb{C}^*$ on $\mathfrak{n}^0$ induces an action of $\mathbb{C}^*$ on $K_{\mathbb{C}} \times \mathfrak{n}^0$ which commutes with the action of N and thus induces an action of $\mathbb{C}^*$ on $(K_{\mathbb{C}} \times \mathfrak{n}^0) /\!\!/ N$. Since $\mathbb{C}^*$ acts on $\mathcal{O}(K_{\mathbb{C}} \times \mathfrak{n}^0)$, and thus on $\mathcal{O}(K_{\mathbb{C}} \times \mathfrak{n}^0)^N$, with only non-negative weights, the sum of the strictly positive weight spaces forms an ideal I in $\mathcal{O}(K_{\mathbb{C}} \times \mathfrak{n}^0)^N$ which defines the fixed point set for the action of $\mathbb{C}^*$ on $(K_{\mathbb{C}} \times \mathfrak{n}^0) /\!\!/ N$. This fixed point set is therefore the affine variety $\operatorname{Spec}(\mathcal{O}(K_{\mathbb{C}} \times \mathfrak{n}^0)^N / I)$, which can be naturally identified with $\operatorname{Spec}((\mathcal{O}(K_{\mathbb{C}} \times \mathfrak{n}^0)^N)^{\mathbb{C}^*})$ and thus with the universal symplectic implosion $K_{\mathbb{C}} /\!\!/ N = \operatorname{Spec}(\mathcal{O}(K_{\mathbb{C}})^N)$.

(5) If $\zeta = (\zeta_1, \zeta_2, \zeta_3) \in \mathfrak{k}^* \otimes \mathbb{R}^3$ let

$$K_{\zeta} = K_{\zeta_1} \cap K_{\zeta_2} \cap K_{\zeta_3}$$

where K_{ζ_j} is the stabiliser of ζ_j under the coadjoint action, and let $\mathcal{N}_\zeta$ be the nilpotent cone in $(\mathfrak{k}_\zeta)^*_{\mathbb{C}}$ which we identify with $(\mathfrak{k}_\zeta)_{\mathbb{C}}$ as usual. By work of Kronheimer [Kro2] there is a $K_\zeta \times T \times SU(2)$-equivariant embedding

$$\mathcal{N}_\zeta \to \mathfrak{k}_\zeta \otimes \mathbb{R}^3$$

whose composition with the projection from $\mathfrak{k}_\zeta \otimes \mathbb{R}^3$ to $(\mathfrak{k}_\zeta)_{\mathbb{C}}$ for any choice of complex structures is the inclusion of the nilpotent cone $\mathcal{N}_\zeta$ in $(\mathfrak{k}_\zeta)_{\mathbb{C}}$. From the discussion in Sect. 3 we expect that for any compact group K the image of the hyperkähler moment map for the action of K on the universal hyperkähler implosion Q should be the K-sweep of

$$\mathfrak{t}_{(\mathrm{hk})} = \{ \zeta + \xi \in \mathfrak{k} \otimes \mathbb{R}^3 : \zeta \in \mathfrak{t} \otimes \mathbb{R}^3 \text{ and } \xi \in \mathcal{N}_\zeta \}$$

and the hyperkähler implosion $X_{\mathrm{hkimpl}} = (X \times Q) /\!/\!/ K$ for any hyperkähler manifold X with a Hamiltonian hyperkähler action of K and hyperkähler moment map $\mu_X : X \to \mathfrak{k} \otimes \mathbb{R}^3$ should be given by

$$X_{\mathrm{hkimpl}} = \mu_X^{-1}(\mathfrak{t}_{(\mathrm{hk})}) / \sim .$$

Here $x \sim y$ if and only if $\mu_X(x) = \zeta + \xi$ and $\mu_X(y) = w(\zeta + \xi')$ for some $\zeta \in \mathfrak{t} \otimes \mathbb{R}^3$, some $\xi, \xi' \in \mathcal{N}_\zeta \subseteq \mathfrak{k} \otimes \mathbb{R}^3$, some w in the Weyl group W of K, identified with a finite subgroup of the normaliser of T in K, and moreover $x = kw^{-1}y$ for some $k \in [K_\zeta, K_\zeta]$.

5 Hyperkähler Implosion for Special Orthogonal and Symplectic Groups

In the case of $K = SU(n)$ the quiver model for the universal symplectic implosion is a symplectic quotient of a flat linear space, so to obtain a quiver model for the universal hyperkähler implosion we could take its cotangent bundle (replacing symplectic with hyperkähler quivers) and the corresponding hyperkähler quotient.

We would like to mimic this construction for the orthogonal and symplectic groups. However we now have the problem that the space of symplectic quivers has a non-flat piece since the top map α_{r-1} has to satisfy the system of quadrics (2) given by $\alpha_{r-1}^t J \alpha_{r-1} = 0$.

If this system of equations cut out a smooth variety we could appeal to a result of Feix [Fei] (see also Kaledin [Kal]) that gives a hyperkähler structure on an open neighbourhood of the zero section of the cotangent bundle of a Kähler manifold with real-analytic metric. In our case, however, the variety defined by (2) is singular. We could of course stratify into smooth varieties by the rank of α_{r-1} and apply Feix's result stratum by stratum, but to obtain a suitable hyperkähler thickening a more global approach is required.

The discussion in Sects. 3 and 4 suggests that we should consider first what the analogue of the hypertoric variety Q_T might be when K is a symplectic or special orthogonal group. As in Sect. 2 let us first consider the case of $K = SO(n)$ when $n = 2r - 1$ is odd.

For the universal symplectic implosion in this case we considered symplectic quivers

$$0 \stackrel{\alpha_0}{\to} \mathbb{C} \stackrel{\alpha_1}{\to} \mathbb{C}^2 \stackrel{\alpha_2}{\to} \cdots \stackrel{\alpha_{r-2}}{\to} \mathbb{C}^{r-1} \stackrel{\alpha_{r-1}}{\to} \mathbb{C}^n \tag{18}$$

and imposed the constraint $\alpha_{r-1}^t J \alpha_{r-1} = 0$; we then took the symplectic quotient by $H_r = \prod_{j=1}^{r-1} SU(j)$ with respect to the standard Kähler structure on $(\mathbb{C}^{j-1})^* \otimes \mathbb{C}^j$ for $j \leqslant r-1$ and the Kähler structure induced by J on $(\mathbb{C}^{r-1})^* \otimes \mathbb{C}^n \cong (\mathbb{C}^n)^{r-1}$. We saw that there is a natural map to this symplectic quotient from the toric variety given by the symplectic quotient of the space of symplectic quivers as above where each map α_j has the form

$$\alpha_j = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ \nu_1^j & 0 & \ldots & 0 \\ 0 & \nu_2^j & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \nu_j^j \end{pmatrix}$$

for $j < r-1$ and

$$\alpha_{r-1} = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 \\ \nu_1^{r-1} & 0 & \ldots & 0 \\ 0 & \nu_2^{r-1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \nu_{r-1}^{r-1} \end{pmatrix}$$

with $\nu_i^j \in \mathbb{C}$ as at (4); notice that a quiver of this form always satisfies the constraint $\alpha_{r-1}^t J \alpha_{r-1} = 0$.

By analogy with this and with the hypertoric variety Q_T described in Sect. 3 for the case when $K = SU(n)$, we expect the hypertoric variety Q_T for $K = SO(n)$ when $n = 2r - 1$ to be closely related to the hyperkähler quotient by the standard maximal torus T_{H_r} of H_r of the flat space $M_T^{SO(n)}$ given by quiver diagrams

$$0 \underset{\beta_o}{\overset{\alpha_o}{\rightleftarrows}} \mathbb{C} \underset{\beta_1}{\overset{\alpha_1}{\rightleftarrows}} \mathbb{C}^2 \underset{\beta_2}{\overset{\alpha_2}{\rightleftarrows}} \cdots \underset{\beta_{r-2}}{\overset{\alpha_{r-2}}{\rightleftarrows}} \mathbb{C}^{r-1} \underset{\beta_{r-1}}{\overset{\alpha_{r-1}}{\rightleftarrows}} \mathbb{C}^n \tag{19}$$

where the maps α_j and β_j have the form

$$\alpha_j = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \nu_1^j & 0 & \dots & 0 \\ 0 & \nu_2^j & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \nu_j^j \end{pmatrix} \quad \text{and} \quad \beta_j = \begin{pmatrix} 0 & \mu_1^j & 0 & \dots & 0 & 0 \\ 0 & 0 & \mu_2^j & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \mu_{j-1}^j & 0 \\ 0 & 0 & 0 & \dots & 0 & \mu_j^j \end{pmatrix}$$

if $j < r-1$ and

$$\alpha_{r-1} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \\ \nu_1^{r-1} & 0 & \dots & 0 \\ 0 & \nu_2^{r-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \nu_{r-1}^{r-1} \end{pmatrix}$$

and

$$\beta_{r-1} = \begin{pmatrix} 0 & \dots & 0 & \mu_1^{r-1} & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & \mu_2^{r-1} & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & \mu_{r-2}^{r-1} & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & \mu_{r-1}^{r-1} \end{pmatrix}$$

for some $\nu_i^j, \mu_i^j \in \mathbb{C} \setminus \{0\}$. Notice that

$$\alpha_{r-1}^t J \alpha_{r-1} = 0 = \beta_{r-1} J \beta_{r-1}^t \tag{20}$$

for any α_{r-1} and β_{r-1} of this form.

Let $M^{SO(n)}$ be the flat hyperkähler space given by arbitrary quiver diagrams of the form (19), where the hyperkähler structure is induced by the standard hyperkähler structure on $(\mathbb{C}^{j-1})^* \otimes \mathbb{C}^j \oplus (\mathbb{C}^j)^* \otimes \mathbb{C}^{j-1} \cong \mathbb{H}^{(j-1)j}$ for $j \leqslant r-1$ and the hyperkähler structure induced by J on $(\mathbb{C}^{r-1})^* \otimes \mathbb{C}^n \oplus (\mathbb{C}^n)^* \otimes \mathbb{C}^{r-1} \cong (\mathbb{H}^n)^{r-1}$. As in the symplectic case discussed in Sect. 2, the restriction to $M_T^{SO(n)}$ of the hyperkähler moment map for the action of H_r coincides with the hyperkähler moment map for the action of T_{H_r} on $M_T^{SO(n)}$. Thus

$$Q_T^{SO(n)} = M_T^{SO(n)} /\!/\!/ T_{H_r}$$

maps naturally to $M^{SO(n)} /\!/\!/ H_r$.

By analogy with the discussion in Sect. 4 we can consider the subset of $M^{SO(n)} /\!/\!/ H_r$ which is the closure of the $K_{\mathbb{C}} = SO(n, \mathbb{C})$-sweep of the image of $Q_T^{SO(n)}$. By (20) this is contained in the closed subset defined by the $K_{\mathbb{C}}$-invariant constraints

$$\alpha_{r-1}^t J \alpha_{r-1} = 0 = \beta_{r-1} J \beta_{r-1}^t,$$

and this closed subset has the dimension expected of the universal hyperkähler implosion. Thus we expect the subset of the hyperkähler quotient $M^{SO(n)} /\!/\!/ H_r$ defined by these constraints to be closely related to the universal hyperkähler implosion for $K = SO(n)$ when $n = 2r - 1$ is odd. Similarly we expect that modifications of this construction as described in Sect. 2 for the universal symplectic implosion will be closely related to the universal hyperkähler implosion for the special orthogonal groups $K = SO(n)$ when n is even and for the symplectic groups. The following example, however, provides a warning against over-optimism here.

Example 5.1. Recall that $SO(3) = SU(2)/\{\pm 1\}$ and that the universal symplectic implosion for $SO(3)$ is $\mathbb{C}^2/\{\pm 1\}$, where $\mathbb{C}^2$ is the universal symplectic implosion for $SU(2)$. Moreover the universal hyperkähler implosion for $SU(2)$ is $\mathbb{H}^2$, given by quiver diagrams of the form

$$\mathbb{C} \underset{\beta}{\overset{\alpha}{\rightleftarrows}} \mathbb{C}^2, \tag{21}$$

(recall that the group H_1 here is trivial, as in Example 2.3, so no quotienting occurs). We thus expect the universal hyperkähler implosion for $SO(3)$ to be $\mathbb{H}^2/\{\pm 1\}$.

We can associate to any quiver (21) the quiver

$$\mathbb{C} \cong \mathrm{Sym}^2(\mathbb{C}) \underset{\beta_1}{\overset{\alpha_1}{\rightleftarrows}} \mathrm{Sym}^2(\mathbb{C}^2) \cong \mathbb{C}^3$$

where $\alpha_1 = \mathrm{Sym}^2(\alpha)$ and $\beta_1 = \mathrm{Sym}^2(\beta)$ are the maps between $\mathrm{Sym}^2(\mathbb{C})$ and $\mathrm{Sym}^2(\mathbb{C}^2)$ induced by α and β. This construction gives us a surjection from $\mathbb{H}^2$ to the subvariety of the space $M^{SO(3)} /\!/\!/ H_1 = M^{SO(3)}$ of quivers

$$\mathbb{C} \underset{\beta_1}{\overset{\alpha_1}{\rightleftarrows}} \mathbb{C}^3$$

satisfying $\alpha_1^t J \alpha_1 = 0 = \beta_1 J \beta_1^t$, but it gives an identification of this subvariety with the quotient of $\mathbb{H}^2$ by $\mathbb{Z}_2 \times \mathbb{Z}_2$, not by $\mathbb{Z}_2 = \{\pm 1\}$.

References

[Bes] A. Besse, *Einstein Manifolds* (Springer, Berlin, 1987)

[Biq] O. Biquard, Sur les équations de Nahm et la structure de Poisson des algèbres de Lie semi-simples complexes. Math. Ann. **304**, 253–276 (1996)

[CG] N. Chriss, V. Ginzburg, *Representation Theory and Complex Geometry* (Birkhauser, Boston, 1997)

[Cox] D. Cox, The homogeneous coordinate ring of a toric variety. J. Algebraic Geom. **4**, 17–50 (1995)

[DKS1] A. Dancer, F. Kirwan, A. Swann, Implosion for hyperkähler manifolds. Compos. Math. **149**, 592–630 (2013)

[DKS2] A. Dancer, F. Kirwan, A. Swann, Implosions and hypertoric geometry. J. Ramanujan Math. Soc. **28A**(special issue in honour of Professor C.S. Seshadri's 80th birthday), 81–122 (2013)

[DKS3] A. Dancer, F. Kirwan, A. Swann, Twistor spaces for hyperkähler implosions. J. Differ. Geom. **97**, 37–77 (2014)

[Fei] B. Feix, Hyperkähler metrics on cotangent bundles. J. Reine Angew. Math. **532**, 33–46 (2001)

[GR] V. Ginzburg, S. Riche, Differential operators on G/U and the affine Grassmannian. J. Inst. Math. J. **14**, 493–575 (2015). arXiv:1306.6754(math.RT)

[GJS] V. Guillemin, L. Jeffrey, R. Sjamaar, *Symplectic implosion*, Transform. Groups **7**, 155–184 (2002)

[HK] Y. Hu, S. Keel, Mori dream spaces and GIT. Mich. Math. J. **48**, 331–348 (2000)

[Kal] D. Kaledin, A canonical hyperkähler metric on the total space of a cotangent bundle, in *Quaternionic Structures in Mathematics and Physics (Rome 1999)* (Univ. Studi. Roma 'La Sapienza', Rome, 1999), pp. 195–230

[Kos] B. Kostant, Lie group representations on polynomial rings. Am. J. Math. **85**, 327–404 (1963)

[Kov] A. Kovalev, Nahm's equations and complex adjoint orbits. Q. J. Math. **47**, 41–58 (1996)

[KS] P. Kobak, A. Swann, Classical nilpotent orbits as hyperkähler quotients. Int. J. Math. **7**, 193–210 (1996)

[KP1] H. Kraft, C. Procesi, Minimal singularities in GL_n. Invent. Math. **62**, 503–515 (1981)

[KP2] H. Kraft, C. Procesi, On the geometry of conjugacy classes in classical groups. Comment. Math. Helv. **57**, 539–602 (1982)

[Kro1] P.B. Kronheimer, A hyperkähler structure on the cotangent bundle of a compact Lie group (1986, preprint), arXiv:math/0409253(DG)

[Kro2] P.B. Kronheimer, Instantons and the geometry of the nilpotent variety. J. Differ. Geom. **32**, 473–490 (1990)

[Kro3] P.B. Kronheimer, A hyper-kählerian structure on coadjoint orbits of a semisimple complex group. J. Lond. Math. Soc. **42**, 193–208 (1990)

[LV] A. Laface, M. Velasco, A survey on Cox rings. Geom. Dedicata. **139**, 269–287 (2009)

[LY] B. Lian, S-T. Yau, Period integrals of CY and general type complete intersections. Invent. Math. **191**, 35–89 (2013)

Kazhdan–Lusztig Conjectures and Shadows of Hodge Theory

Ben Elias and Geordie Williamson

Abstract We give an informal introduction to the authors' work on some conjectures of Kazhdan and Lusztig, building on work of Soergel and de Cataldo–Migliorini. This article is an expanded version of a lecture given by the second author at the Arbeitstagung in memory of Frederich Hirzebruch.

1 Introduction

It was a surprise and honour to be able to speak about our recent work at the Arbeitstagung in memory of Hirzebruch. These feelings are heightened by the fact that the decisive moments in the development of our joint work occurred at the Max-Planck-Institute in Bonn, which owes its very existence to Hirzebruch. In the following introduction we have tried to emphasize the aspects of our work which we believe Hirzebruch would have most enjoyed: compact Lie groups and the topology of their homogenous spaces; characteristic classes; Hodge theory; and more generally the remarkable topological properties of projective algebraic varieties.

Let G be a connected compact Lie group and T a maximal torus. A fundamental object in mathematics is the flag manifold G/T. We briefly recall Borel's beautiful and canonical description of its cohomology. Given a character $\lambda : T \to \mathbb{C}^*$ we can form the line bundle

$$L_\lambda := G \times_T \mathbb{C}$$

on G/T, defined as the quotient of $G \times \mathbb{C}$ by T-action given by $t \cdot (g, x) := (gt^{-1}, \lambda(t)x)$. Taking the Chern class of L_λ yields a homomorphism

$$X(T) \to H^2(G/T) : \lambda \mapsto c_1(L_\lambda).$$

B. Elias
Department of Mathematics, University of Oregon, Eugene, OR, USA
e-mail: belias@uoregon.edu

G. Williamson (✉)
Max-Planck-Institut für Mathematik, Vivatsgasse 7, 53111 Bonn, Germany
e-mail: geordie@mpim-bonn.mpg.de

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_5

from the lattice of characters to the second cohomology of G/T. If we identify $X(T) \otimes_{\mathbb{Z}} \mathbb{R} = (\mathrm{Lie}\, T)^*$ via the differential and extend multiplicatively we get a morphism of graded algebras

$$R := S((\mathrm{Lie}\, T)^*) \to H^\bullet(G/T; \mathbb{R}).$$

called the *Borel homomorphism*. (We let R denote the symmetric algebra on the dual of $\mathrm{Lie}\, T$.) Borel showed that his homomorphism is surjective and identified its kernel with the ideal generated by W-invariant polynomials of positive degree. Here $W = N_G(T)/T$ denotes the Weyl group of G which acts on T by conjugation, hence on $\mathrm{Lie}\, T$ and hence on R.

For example, let $G = U(n)$ be the unitary group, and T the subgroup of diagonal matrices ($\cong (S^1)^n$). Then the coordinate functions give an identification $R = \mathbb{R}[x_1, \dots, x_n]$, and W is the symmetric group on n-letters acting on R via permutation of variables. The Borel homomorphism gives an identification

$$\mathbb{R}[x_1, \dots, x_n]/\langle e_i \mid 1 \le i \le n \rangle = H^\bullet(G/T; \mathbb{R})$$

where e_i denotes the ith elementary symmetric polynomial in $x_1, \dots, x_n$.

Let $G_{\mathbb{C}}$ denote the complexification of G and choose a Borel subgroup B containing the complexification of T. (For example, if $G = U(n)$, then $G_{\mathbb{C}} = GL_n(\mathbb{C})$ and we could take B to be the subgroup of upper-triangular matrices.) A fundamental fact is that the natural map

$$G/T \to G_{\mathbb{C}}/B$$

is a diffeomorphism, and $G_{\mathbb{C}}/B$ is a projective algebraic variety.

For example, if $G = SU(2) \cong S^3$, then $G/T = S^2$ is the base of the Hopf fibration, and the above diffeomorphism is $S^2 \xrightarrow{\sim} \mathbb{P}^1\mathbb{C}$. More generally for $G = U(n)$ the above diffeomorphism can be seen as an instance of Gram-Schmidt orthogonalization. Fix a Hermitian form on $\mathbb{C}^n$. Then $G_{\mathbb{C}}/B$ parametrizes complete flags on $\mathbb{C}^n$, while G/T parametrizes collections of n ordered orthogonal complex lines. These spaces are clearly isomorphic.

The fact that $G/T = G_{\mathbb{C}}/B$ is a projective algebraic variety means that its cohomology satisfies a number of deep theorems from complex algebraic geometry. Set $H = H^\bullet(G_{\mathbb{C}}/B; \mathbb{R})$ and let N denote the complex dimension of $G_{\mathbb{C}}/B$. For us the following two results (the "shadows of Hodge theory" of the title) will be of fundamental importance.

Theorem 1.1 (Hard Lefschetz Theorem). *Let $\lambda \in H^2$ denote the Chern class of an ample line bundle on $G_{\mathbb{C}}/B$ [i.e. $\lambda \in (\mathrm{Lie}\, T)^*$ is a "dominant weight", see (3)]. Then for all $0 \le i \le N$ multiplication by λ^{N-i} gives an isomorphism:*

$$\lambda^{N-i} : H^i \xrightarrow{\sim} H^{2N-i}.$$

Because G/T is a compact manifold, Poincaré duality states that H^i and H^{2N-i} are non-degenerately paired by the Poincaré pairing $\langle -,-\rangle_{\text{Poinc}}$. On the other hand, after fixing λ as above the hard Lefschetz theorem gives us a way of identifying H^i and H^{2N-i}. The upshot is that for $0 \le i \le N$ we obtain a non-degenerate *Lefschetz form*:

$$H^i \times H^i \to \mathbb{R}$$
$$(\alpha, \beta) \mapsto \langle \alpha, \lambda^{N-i}\beta\rangle_{\text{Poinc}}.$$

On the middle dimensional cohomology the Lefschetz form is just the Poincaré pairing. This is the only Lefschetz form which does not depend on the choice of ample class λ.

Theorem 1.2 (Hodge-Riemann Bilinear Relations). *For $0 \le i \le N$ the restriction of the Lefschetz form to $P^i := \ker(\lambda^{N-i+1}) \subset H^i$ is $(-1)^{i/2}$-definite.*

Some comments are in order:

(1) The odd cohomology of G/T vanishes as can be seen, for example, from the surjectivity of the Borel homomorphism. Hence the sign $(-1)^{i/2}$ makes sense.
(2) For an arbitrary smooth projective algebraic variety the Hodge-Riemann bilinear relations are more complicated, involving the Hodge decomposition and a Hermitian form on the complex cohomology groups. However, the cohomology of the flag variety is always in (p, p)-type, so that we may use the simpler formulation above.
(3) We will not make it explicit, but the Hodge-Riemann bilinear relations give formulas for the signatures of all Lefschetz forms in terms of the graded dimension of H.

We now come to the punchline of this survey. The hard Lefschetz theorem and Hodge-Riemann bilinear relations for $H^\bullet(G/B;\mathbb{R})$ are deep consequences of Hodge theory. On the other hand, we have seen that the Borel homomorphism gives us an elementary description of $H^\bullet(G/B;\mathbb{R})$ in terms of commutative algebra and invariant theory. Can one establish the hard Lefschetz theorem and Hodge-Riemann bilinear relations for $H^\bullet(G/B;\mathbb{R})$ algebraically? A crucial motivation for this question is the fact that $H^\bullet(G/B;\mathbb{R})$ has various algebraic cousins (described in Sect. 5) for which no geometric description is known. Remarkably, these cousins still satisfy analogs of Theorems 1.1 and 1.2. Establishing these Hodge-theoretic properties algebraically is the cornerstone of the authors' approach to conjectures of Kazhdan–Lusztig and Soergel.

The structure of this (very informal) survey is as follows. In Sect. 2 we give a lightning introduction to intersection cohomology, which provides an improved cohomology theory for singular algebraic varieties. In Sect. 3 we discuss Schubert varieties, certain (usually singular) subvarieties of the flag variety which play an important role in representation theory. We also discuss Bott–Samelson resolutions of Schubert varieties. In Sect. 4 we discuss Soergel modules. The point is that one can give a purely algebraic/combinatorial description of the intersection cohomology of Schubert varieties, which only depends on the underlying Weyl group. In Sect. 5 we discuss Soergel modules for arbitrary Coxeter groups, which (currently) have no geometric interpretation. We also state our main theorem that these modules satisfy the "shadows of Hodge theory". Finally, in Sect. 6 we discuss the amusing example of the coinvariant ring of a finite dihedral group.

2 Intersection Cohomology and the Decomposition Theorem

Poincaré duality, the hard Lefschetz theorem and Hodge-Riemann bilinear relations hold for the cohomology of any smooth projective variety. The statements of these results usually fail for singular varieties. However, in the 1970s Goresky and MacPherson invented intersection cohomology [GM80, GM83] and it was later proven that the analogues of these theorems hold for intersection cohomology. In this section we will try to give the vaguest of vague ideas as to what is going on, and hopefully convince the reader to go and read more. (The authors' favourite introduction to the theory is [dM09] whose emphasis agrees largely with that of this survey.[1] More information is contained in [Bor94, Rie04, Ara06] with the bible being [BBD82]. To stay motivated, Kleiman's excellent history of the subject [Kle07] is a must.)

Intersection cohomology associates to any complex variety X its "intersection cohomology groups" $IH^\bullet(X)$ (throughout this article we always take coefficients in $\mathbb{R}$, however, there are versions of the theory with $\mathbb{Q}$ and $\mathbb{Z}$-coefficients). Here are some basic properties of intersection cohomology:

(1) $IH^\bullet(X)$ is a graded vector space, concentrated in degrees between 0 and $2N$, where N is the complex dimension of X;
(2) if X is smooth, then $IH^\bullet(X) = H^\bullet(X)$;
(3) if X is projective, then $IH^\bullet(X)$ is equipped with a non-degenerate Poincaré pairing $\langle -, -\rangle_{\text{Poinc}}$, which is the usual Poincaré pairing for X smooth.

However we caution the reader that:

(1) the assignment $X \mapsto IH^\bullet(X)$ is not functorial: in general a morphism $f : X \to Y$ does not induce a pull-back map on intersection cohomology;
(2) $IH^\bullet(X)$ is not a ring, but rather a module over the cohomology ring $H^\bullet(X)$.

(These two "failings" become less worrying when one interprets intersection cohomology in the language of constructible sheaves.) Finally, we come to the two key properties that will concern us in this article. We assume that X is a projective variety (not necessarily smooth):

[1] Due, no doubt, to the influence which their work has had on the authors.

(1) multiplication by the first Chern class of an ample line bundle on $IH^\bullet(X)$ satisfies the hard Lefschetz theorem;
(2) the groups $IH^\bullet(X)$ satisfy the Hodge-Riemann bilinear relations.

(To make sense of this second statement, one needs to know that $IH^\bullet(X)$ has a Hodge decomposition. This is true, but we will not discuss it. Below, we will only consider varieties whose Hodge decomposition only involves components of type (p,p) and so the naive formulation of the Hodge-Riemann bilinear relations in the form of Theorem 1.2 will be sufficient.)

Example 2.1. Consider the Grassmannian $\mathrm{Gr}(2,4)$ of planes in $\mathbb{C}^4$. It is a smooth projective algebraic variety of complex dimension 4. Let $0 \subset \mathbb{C} \subset \mathbb{C}^2 \subset \mathbb{C}^3 \subset \mathbb{C}^4$ denote the standard coordinate flag on $\mathbb{C}^4$. For any sequence of natural numbers $\underline{a} := (0 = a_0 \le a_1 \le a_2 \le a_3 \le a_4 = 2)$ satisfying $a_i \le a_{i+1} \le a_i + 1$, consider the subvariety

$$C_{\underline{a}} := \{V \in \mathrm{Gr}(2,4) \mid \dim(V \cap \mathbb{C}^i) = a_i\}.$$

It is not difficult (by writing down charts for the Grassmannian) to see that each $C_{\underline{a}}$ is isomorphic to $\mathbb{C}^{d(\underline{a})}$ where $d(\underline{a}) = 7 - \sum_{i=0}^4 a_i$. Hence $\mathrm{Gr}(2,4)$ has a cell-decomposition with cells of real dimension $0,2,4,4,6,8$. The cohomology $H^\bullet(\mathrm{Gr}(2,4))$ is as follows:

0	1	2	3	4	5	6	7	8
$\mathbb{R}$	0	$\mathbb{R}$	0	$\mathbb{R}^2$	0	$\mathbb{R}$	0	$\mathbb{R}$

It is an easy exercise to use Schubert calculus (see, e.g., [Hil82, III.3], which also discusses $\mathrm{Gr}(2,4)$ in more detail) to check the hard Lefschetz theorem and Hodge-Riemann bilinear relations by hand.

Now consider the subvariety

$$X := \{V \in \mathrm{Gr}(2,4) \mid \dim(V \cap \mathbb{C}^2) \ge 1\}.$$

Then X coincides with the closure of the cell $C_{0\le 0\le 1\le 1\le 2} \subset \mathrm{Gr}(2,4)$ (and thus is an example of a "Schubert variety", as we will discuss in the next section). Hence X has real dimension 6 and has a cell-decomposition with cells of dimension $(0,2,4,4,6)$. Its cohomology is as follows:

0	1	2	3	4	5	6
$\mathbb{R}$	0	$\mathbb{R}$	0	$\mathbb{R}^2$	0	$\mathbb{R}$

We conclude that X cannot satisfy Poincaré duality or the hard Lefschetz theorem. In particular X must be singular. In fact, X has a unique singular point $V_0 = \mathbb{C}^2$. We will see below that the intersection cohomology $IH^\bullet(X)$ is as follows:

0	1	2	3	4	5	6
$\mathbb{R}$	0	$\mathbb{R}^2$	0	$\mathbb{R}^2$	0	$\mathbb{R}$

So in this example $IH^\bullet(X)$ seems to fit the bill (at least on the level of Betti numbers) of rescuing Poincaré duality and the hard Lefschetz theorem in a "minimal" way.

Probably the most fundamental theorem about intersection cohomology is the decomposition theorem. In its simplest form it says the following:

Theorem 2.2 (Decomposition Theorem [BBD82, Sai89, dCM02, dCM05]). *Let $f : \widetilde{X} \to X$ be a resolution, i.e., $\widetilde{X}$ is smooth and f is a projective birational morphism of algebraic varieties. Then $IH^\bullet(X)$ is a direct summand of $H^\bullet(\widetilde{X})$, as modules over $H^\bullet(X)$.*

The decomposition theorem provides an invaluable tool for calculating intersection cohomology, which is otherwise a very difficult task.

Example 2.3. In Example 2.1 we discussed the variety

$$X := \{V \in \mathrm{Gr}(2,4) \mid \dim(V \cap \mathbb{C}^2) \geq 1\}$$

which is projective with unique singular point $V_0 = \mathbb{C}^2$. Now X has a natural resolution $f : \widetilde{X} \to X$ where

$$\widetilde{X} = \{(V,W) \in \mathrm{Gr}(2,4) \times \mathbb{P}(\mathbb{C}^2) \mid W \subset V \cap \mathbb{C}^2\}$$

and $f(V,W) = V$. Clearly f is an isomorphism over $X \setminus \{V_0\}$ and has fibre $\mathbb{P}^1 = \mathbb{P}(\mathbb{C}^2)$ over the singular point V_0. Also, the projection $(V,W) \mapsto W$ realizes $\widetilde{X}$ as a $\mathbb{P}^2$-bundle over $\mathbb{P}^1$. In particular, $\widetilde{X}$ is smooth and its cohomology is as follows:

0	1	2	3	4	5	6
$\mathbb{R}$	0	$\mathbb{R}^2$	0	$\mathbb{R}^2$	0	$\mathbb{R}$

We conclude by the decomposition theorem that $IH^\bullet(X)$ is a summand of $H^\bullet(\widetilde{X})$. In this case one has equality: $IH^\bullet(X) = H^\bullet(\widetilde{X})$. One can see this directly as follows: first one checks that the pull-back map $H^i(X) \to H^i(\widetilde{X})$ is injective. Now, because $IH^\bullet(X)$ is an $H^\bullet(X)$-stable summand of $H^\bullet(\widetilde{X})$ containing $\mathbb{R} = H^0(\widetilde{X})$ we conclude that $IH^i(X) = H^i(\widetilde{X})$ for $i \neq 2$. Finally, we must have $IH^2(X) = H^2(\widetilde{X})$ because $IH^\bullet(X)$ satisfies Poincaré duality.

Let us now discuss the hard Lefschetz theorem and Hodge-Riemann bilinear relations for $IH^\bullet(X)$. Let λ be the class of an ample line bundle on X. Because $IH^\bullet(X) = H^\bullet(\widetilde{X})$ in this example, the action of λ on $IH^\bullet(X)$ is identified with the action of $f^*\lambda$ on $H^\bullet(\widetilde{X})$. We would like to know that $f^*\lambda$ acting on $H^\bullet(\widetilde{X})$ satisfies the hard Lefschetz theorem and Hodge-Riemann bilinear relations *even though $f^*\lambda$ is not an ample class on $\widetilde{X}$*. This simple observation is the starting point for beautiful work of de Cataldo and Migliorini [dCM02, dCM05], who give a Hodge-theoretic proof of the decomposition theorem.

3 Schubert Varieties and Bott–Samelson Resolutions

Recall our connected compact Lie group G, its complexification $G_{\mathbb{C}}$, the maximal torus $T \subset G$ and the Borel subgroup $T \subset B \subset G_{\mathbb{C}}$. To (G, T) we may associate a root system $\Phi \subset (\text{Lie}\, T)^*$. Our choice of Borel subgroup is equivalent to a choice of simple roots $\Delta \subset \Phi$. As we discussed in the introduction, the Weyl group $W = N_G(T)/T$ acts on Lie T as a reflection group. The choice of simple roots $\Delta \subset \Phi$ gives a choice of *simple reflections* $S \subset W$. These simple reflections generate W and with respect to these generators W admits a *Coxeter presentation*:

$$W = \langle s \in S \,|s^2 = \text{id}, (st)^{m_{st}} = \text{id}\rangle$$

where $m_{st} \in \{2, 3, 4, 6\}$ can be read off the Dynkin diagram of G. Given $w \in W$ a *reduced expression* for w is an expression $w = s_1 \dots s_m$ with $s_i \in S$, having shortest length amongst all such expressions. The *length* $\ell(w)$ of w is the length of a reduced expression. The Weyl group W is finite, with a unique longest element w_0.

From now on we will work with the flag variety $G_{\mathbb{C}}/B$ in its incarnation as a projective algebraic variety. It is an important fact (the "Bruhat decomposition") that B has finitely many orbits on $G_{\mathbb{C}}/B$ which are parametrized by the Weyl group W. In formulas we write

$$G_{\mathbb{C}}/B = \bigsqcup_{w \in W} B \cdot wB/B$$

Each B-orbit $B \cdot wB/B$ is isomorphic to an affine space and its closure

$$X_w := \overline{B \cdot wB/B}$$

is a projective variety called a *Schubert variety*. It is of complex dimension $\ell(w)$. The two extreme cases are $X_{\text{id}} = B/B$, a point, and $X_{w_0} = G_{\mathbb{C}}/B$, the full flag variety.

More generally, given any subset $I \subset S$ we have a parabolic subgroup $B \subset P_I \subset G$ generated by B and (any choice of representatives of) the subset I. The quotient G/P_I is also a projective algebraic variety (called a *partial flag variety*) and the Bruhat decomposition takes the form

$$G/P_I := \bigsqcup_{w \in W^I} B \cdot wB/P_I$$

where W^I denotes a set of minimal length representatives for the cosets W/W_I. Again, the Schubert varieties are the closures $X_w^I := \overline{B \cdot wB/P_I} \subset G/P_I$, which are projective algebraic varieties of dimension $\ell(w)$.

Example 3.1. We discussed the more general setting of G/P_I to make contact with the Grassmannian in Example 2.1. Indeed, $\text{Gr}(2, 4) \cong GL_4(\mathbb{C})/P$ where P is the stabilizer of the fixed coordinate subspace $\mathbb{C}^2 \subset \mathbb{C}^4$. If B denotes the stabilizer of the coordinate flag $0 \subset \mathbb{C}^1 \subset \mathbb{C}^2 \subset \mathbb{C}^3 \subset \mathbb{C}^4$ (the upper-triangular matrices), then the cells $C_{\underline{a}}$ of Example 2.1 are B-orbits on $\mathrm{Gr}(2,4)$. Hence our X is an example of a singular Schubert variety.

Schubert varieties are rarely smooth. We now discuss how to construct resolutions. We will focus on Schubert varieties in the full flag variety, although similar constructions work for Schubert varieties in partial flag varieties. Choose $w \in W$ and fix a reduced expression $w = s_1 s_2 \dots s_m$. For any $1 \le i \le m$ let us alter our notation and write P_i for $P_{\{s_i\}} = \overline{Bs_iB}$, a (minimal) parabolic subgroup associated with the reflection s_i. Consider the space

$$\mathrm{BS}(s_1, \dots, s_m) := P_1 \times^B P_2 \times^B \cdots \times^B P_m/B.$$

The notation $\times^B$ indicates that $\mathrm{BS}(s_1, \dots, s_m)$ is the quotient of $P_1 \times P_2 \times \cdots \times P_m$ by the action of B^m via

$$(b_1, b_2, \dots, b_m) \cdot (p_1, \dots, p_m) = (p_1 b_1^{-1}, b_1 p_2 b_2^{-1}, \dots, b_{m-1} p_m b_m^{-1}).$$

Then $\mathrm{BS}(s_1, \dots, s_m)$ is a smooth projective *Bott–Samelson variety* and the multiplication map $P_1 \times \cdots \times P_m \to G$ induces a morphism

$$f : \mathrm{BS}(s_1, \dots, s_m) \to X_w$$

which is a resolution of X_w. (See [Dem74, Han73] and [Bri12, Sect. 2] for further discussion and applications of Bott–Samelson resolutions. The name Bott–Samelson resolution comes from [BS58] where related spaces are considered in the context of loop spaces of compact Lie groups.)

Example 3.2. If $G_{\mathbb{C}} = GL_n$, Bott–Samelson resolutions admit a more explicit description. Recall that GL_n/B is the variety of flags $V_\bullet = (0 = V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_n = \mathbb{C}^n)$ with $\dim V_i = i$. We identify W with the symmetric group S_n and S with the set of simple transpositions $\{s_i = (i, i+1) \mid 1 \le i \le n-1\}$. Given a reduced expression $s_{i_1} \dots s_{i_m}$ for $w \in W$ consider the variety $\widetilde{\mathrm{BS}}(s_{i_1}, \dots, s_{i_m})$ of all m-tuples of flags $(V_\bullet^a)_{0 \le a \le m}$ such that:

(1) $V_\bullet^0$ is the coordinate flag $V_\bullet^{\mathrm{std}} = (0 \subset \mathbb{C}^1 \subset \cdots \subset \mathbb{C}^n)$;
(2) for all $1 \le a \le m$, $V_j^a = V_j^{a-1}$ for $j \neq i_a$.

That is, $\widetilde{\mathrm{BS}}(s_{i_1}, \dots, s_{i_m})$ is the variety of sequences of $m+1$ flags which begin at the coordinate flag, and where, in passing from the $(j-1)$st to the jth step, we are only allowed to change the i_jth dimensional subspace.

Let $p_0 = 1$. Then the map

$$(p_1, \dots, p_m) \mapsto (p_0 \dots p_a V_\bullet^{\mathrm{std}})_{a=0}^m$$

gives an isomorphism $\mathrm{BS}(s_1, \dots, s_m) \to \widetilde{\mathrm{BS}}(s_1, \dots, s_m)$. Under this isomorphism the map f becomes the projection to the final flag: $f((V_\bullet^a)_{a=1}^m) = V_\bullet^m$.

4 Soergel Modules and Intersection Cohomology

In a landmark paper [Soe90], Soergel explained how to calculate the intersection cohomology of Schubert varieties in a purely algebraic way. Though much less explicit, one way of viewing this result is as a generalization of Borel's description of the cohomology of the flag variety.

The idea is as follows. In the last section we discussed the Bott–Samelson resolutions of Schubert varieties

$$f : \mathrm{BS}(s_1, \dots, s_m) \to X_w \subset G_{\mathbb{C}}/B$$

where $w = s_1 \dots s_m$ is a reduced expression for w. By the decomposition theorem $IH^\bullet(X_w)$, the intersection cohomology of the Schubert variety $X_w \subset G_{\mathbb{C}}/B$, is a summand of $H^\bullet(\mathrm{BS}(s_1, \dots, s_m))$. Moreover, we have pull-back maps

$$H^\bullet(G_{\mathbb{C}}/B) \twoheadrightarrow H^\bullet(X_w) \to H^\bullet(\mathrm{BS}(s_1, \dots, s_m))$$

and $IH^\bullet(X_w)$ is even a summand of $H^\bullet(\mathrm{BS}(s_1, \dots, s_m))$ as an $H^\bullet(G_{\mathbb{C}}/B)$-module. (The surjectivity of the restriction map $H^\bullet(G_{\mathbb{C}}/B) \twoheadrightarrow H^\bullet(X_w)$ follows because both spaces have compatible cell-decompositions.) Remarkably, this algebraic structure already determines the summand $IH^\bullet(X_w)$ (see [Soe90, Erweiterungssatz]):

Theorem 4.1 (Soergel). *Let $w = s_1 \dots s_m$ denote a reduced expression for w as above. Consider $H^\bullet(\mathrm{BS}(s_1, \dots, s_m))$ as a $H^\bullet(G_{\mathbb{C}}/B)$-module. Then $IH(X_w)$ may be described as the indecomposable graded $H^\bullet(G_{\mathbb{C}}/B)$-module direct summand with non-trivial degree zero part.*

A word of caution: The realization of $IH^\bullet(X_w)$ inside $H^\bullet(\mathrm{BS}(s_1, \dots, s_m))$ is not canonical in general. We can certainly decompose $H^\bullet(\mathrm{BS}(s_1, \dots, s_m))$ into graded indecomposable $H^\bullet(G_{\mathbb{C}}/B)$-modules. Although this decomposition is not canonical, the Krull–Schmidt theorem ensures that the isomorphism type and multiplicities of indecomposable summands do not depend on the chosen decomposition. The above theorem states that, for any such decomposition, the unique indecomposable module with non-trivial degree zero part is isomorphic to $IH^\bullet(X_w)$ (as an $H^\bullet(G_{\mathbb{C}}/B)$-module).

We now explain (following Soergel) how one may give an algebraic description of all players in the above theorem. Recall that $R = S((\mathrm{Lie}\, T)^*)$ denotes the symmetric algebra on the dual of $\mathrm{Lie}\, T$, graded so that $(\mathrm{Lie}\, T)^*$ has degree 2. The Weyl group W acts on R, and for any simple reflection $s \in S$ we denote by R^s the invariants under s. It is not difficult to see that R is a free graded module of rank 2 over R^s with basis $\{1, \alpha_s\}$, where α_s is the simple root associated with $s \in S$. (In essence this is the high-school fact that any polynomial can be written as the sum of its even and odd parts.)

The starting point is the following observation:

Proposition 4.2 (Soergel). *One has an isomorphism of graded algebras*

$$H^\bullet(\mathrm{BS}(s_1,\dots,s_m)) = R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} \dots R \otimes_{R^{s_m}} R \otimes_R \mathbb{R}$$

where the final term is an R-algebra via $\mathbb{R} \cong R/R^{>0}$.

For example, for any $s \in S$ we have $\mathrm{BS}(s) = P_s/B \cong \mathbb{P}^1$ and $R \otimes_{R^s} R \otimes_R \mathbb{R} = R \otimes_{R^s} \mathbb{R}$ is 2-dimensional, with graded basis $\{1 \otimes 1, \alpha_s \otimes 1\}$ of degrees 0 and 2. More generally, one can show that

$$R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} \dots R \otimes_{R^{s_m}} R \otimes_R \mathbb{R} = R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} \dots R \otimes_{R^{s_m}} \mathbb{R}$$

has graded basis $\alpha_{s_1}^{\varepsilon_1} \otimes \alpha_{s_2}^{\varepsilon_2} \otimes \dots \otimes \alpha_{s_m}^{\varepsilon_m} \otimes 1$ where $(\varepsilon_a)_{a=1}^m$ is any tuple of zeroes and ones. In particular, its Poincaré polynomial is $(1+q^2)^m$.

Recall that in the introduction we described the Borel isomorphism:

$$H^\bullet(G/B) \cong R/(R_+^W).$$

Notice that left multiplication by any invariant polynomial of positive degree acts as zero on

$$H^\bullet(\mathrm{BS}(s_1,\dots,s_m)) = R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} \dots R \otimes_{R^{s_m}} R \otimes_R \mathbb{R}.$$

We conclude that $R \otimes_{R^{s_1}} \dots R \otimes_{R^{s_m}} R \otimes_R \mathbb{R}$ is a module over $R/(R_+^W)$. Geometrically, this corresponds to the pull-back map on cohomology

$$H^\bullet(G_{\mathbb{C}}/B) \to H^\bullet(\mathrm{BS}(s_1,\dots,s_m))$$

discussed above.

We can now reformulate Theorem 4.1 algebraically as follows:

Theorem 4.3 (Soergel [Soe90]). *Let D_w be any indecomposable $R/(R_+^W)$-module direct summand of*

$$H^\bullet(\mathrm{BS}(s_1,\dots,s_m)) = R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} \dots R \otimes_{R^{s_m}} R \otimes_R \mathbb{R}$$

containing the element $1 \otimes 1 \otimes \dots \otimes 1$, where $w = s_1 \dots s_m$ is a reduced expression for w. Then D_w is well-defined up to isomorphism (i.e. does not depend on the choice of reduced expression) and $D_w \cong IH^\bullet(X_w)$.

The modules $\{D_w \mid w \in W\}$ are the (indecomposable) *Soergel modules*.

Example 4.4. We consider the case of $G = GL_3(\mathbb{C})$ in which case

$$W = S_3 = \{\mathrm{id}, s_1, s_2, s_1s_2, s_2s_1, s_1s_2s_1\}$$

(we use the conventions of Example 3.2). In this case it turns out that all Schubert varieties are smooth. Also, if $\ell(w) \leq 2$, then any Bott–Samelson resolution is an isomorphism. We conclude

$$D_{\mathrm{id}} = \mathbb{R}$$

$$D_{s_1} = H^\bullet(\mathrm{BS}(s_1)) = R \otimes_{R^{s_1}} \mathbb{R} \qquad D_{s_2} = R \otimes_{R^{s_2}} \mathbb{R}$$

$$D_{s_1 s_2} = H^\bullet(\mathrm{BS}(s_1, s_2)) = R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} \mathbb{R} \qquad D_{s_2 s_1} = R \otimes_{R^{s_2}} R \otimes_{R^{s_1}} \mathbb{R}$$

(A pleasant exercise for the reader is to verify that in all these examples above D_x is a cyclic (hence indecomposable) module over R. This is not usually the case, and is related to the (rational) smoothness of the Schubert varieties in question.)

The element $w_0 = s_1 s_2 s_1$ is more interesting. In this case the Bott–Samelson resolution

$$\mathrm{BS}(s_1, s_2, s_1) \to X_{w_0} = G/B$$

is not an isomorphism. As previously discussed, the Poincaré polynomial of

$$H^\bullet(\mathrm{BS}(s_1, s_2, s_1)) = R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} R \otimes_{R^{s_1}} \mathbb{R} \tag{1}$$

is $(1 + q^2)^3$ whereas the Poincaré polynomial of

$$IH^\bullet(X_{w_0}) = H^\bullet(G/B) = R/(R_+^W) \tag{2}$$

is $(1 + q^2)(1 + q^2 + q^4)$. In this case the reader may verify that (2) is a summand of (1). In fact one has an isomorphism of graded $R/(R_+^W)$-modules:

$$R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} R \otimes_{R^{s_1}} \mathbb{R} = R/(R_+^W) \oplus (R \otimes_{R^s} \mathbb{R}(-2)).$$

Here $R \otimes_{R^s} \mathbb{R}(-2)$ denotes the shift of $R \otimes_{R^s} \mathbb{R}$ in the grading such that its generator $1 \otimes 1$ occurs in degree 2. This extra summand can be embedded into (1) via the map which sends

$$f \otimes 1 \mapsto f \otimes \alpha_{s_2} \otimes 1 \otimes 1 + f \otimes 1 \otimes \alpha_{s_2} \otimes 1$$

for $f \in R$.

Example 4.5. If w_0 denotes the longest element of W, then $X_{w_0} = G_{\mathbb{C}}/B$, the (smooth) flag variety of G. In particular

$$IH^\bullet(X_{w_0}) = H^\bullet(G_{\mathbb{C}}/B) = R/(R_+^W)$$

by the Borel isomorphism. Theorem 4.3 asserts that $R/(R_+^W)$ occurs as a direct summand of

$$R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} \otimes \cdots \otimes_{R^{s_m}} \mathbb{R}$$

for any reduced expression $w_0 = s_1 \dots s_m$. This is by no means obvious! We have seen an instance of this in the previous example.

Remark 4.6. In this section we could have worked in the category of graded R-modules, rather than the category of graded $R/(R_+^W)$-modules, and it would change nothing. All the R-modules in question will factor through $R/(R_+^W)$. In the next section, we will work with R-modules instead.

We now discuss hard Lefschetz and the Hodge-Riemann bilinear relations. Recall that our Borel subgroup $B \subset G_{\mathbb{C}}$ determines a set of simple roots $\Delta \subset \Phi \subset (\mathrm{Lie}\, T)^*$ and simple coroots $\Delta^\vee \subset \Phi^\vee \subset \mathrm{Lie}\, T$. Under the isomorphism

$$H^2(G_{\mathbb{C}}/B) \cong (\mathrm{Lie}\, T)^*$$

the ample cone (i.e. the $\mathbb{R}_{>0}$-stable subset of $H^2(G_{\mathbb{C}}/B)$ generated by Chern classes of ample line bundles on $G_{\mathbb{C}}/B$) is the cone of dominant weights for Lie T:

$$(\mathrm{Lie}\, T)_+^* := \{\lambda \in (\mathrm{Lie}\, T)^* \mid \langle \lambda, \alpha^\vee \rangle > 0 \text{ for all } \alpha^\vee \in \Delta^\vee\}. \tag{3}$$

The hard Lefschetz theorem then asserts that left multiplication by any $\lambda \in (\mathrm{Lie}\, T)_+^*$ satisfies the hard Lefschetz theorem on $D_w = IH^\bullet(X_w)$. That is, for all $i \geq 0$, multiplication by λ^i induces an isomorphism

$$\lambda^i : D_w^{\ell(w)-i} \xrightarrow{\sim} D_w^{\ell(w)+i}.$$

To discuss the Hodge-Riemann relations we need to make the Poincaré pairing $\langle -, - \rangle_{\mathrm{Poinc}}$ explicit for D_w. We first discuss the Poincaré form on $H^\bullet(\mathrm{BS}(s_1, \dots, s_m))$. Recall that for any oriented manifold M the Poincaré form in de Rham cohomology is given by

$$\langle \alpha, \beta \rangle = \int_M \alpha \wedge \beta.$$

We imitate this algebraically as follows. By the discussion after Proposition 4.2, the degree $2m$ component of

$$H^\bullet(\mathrm{BS}(s_1, \dots, s_m)) = R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} \dots R \otimes_{R^{s_m}} \mathbb{R}$$

is one-dimensional and is spanned by the vector $c_{\mathrm{top}} := \alpha_{s_1} \otimes \alpha_{s_2} \otimes \cdots \otimes \alpha_{s_m} \otimes 1$. We can define a bilinear form $\langle -, - \rangle$ on $R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} \dots R \otimes_{R^{s_m}} \mathbb{R}$ via

$$\langle f, g \rangle = \mathrm{Tr}(fg)$$

where fg denotes the term-wise multiplication, and Tr is the functional which returns the coefficient of c_{top}. Then $\langle -, - \rangle$ is a non-degenerate symmetric form which agrees up to a positive scalar with the intersection form on $H^\bullet(\mathrm{BS}(s_1, \dots, s_m))$.

Now recall that D_w is obtained as summand of $R \otimes_{R^{s_1}} R \otimes_{R^{s_2}} \dots R \otimes_{R^{s_m}} \mathbb{R}$, for a reduced expression of w. Fixing such an inclusion we obtain a form on D_w via restriction of the form $\langle -, - \rangle$. In fact, this form is well-defined (i.e. depends neither on the choice of reduced expression nor embedding) up to a positive scalar. One can show that this form agrees with the Poincaré pairing on $D_w = IH^\bullet(X_w)$ up to a positive scalar. The Hodge-Riemann bilinear relations then hold for D_w with respect to this form and left multiplication by any $\lambda \in (\text{Lie}\, T)^*_+$.

5 Soergel Modules for Arbitrary Coxeter Systems

Now let (W, S) denote an arbitrary Coxeter system. That is, W is a group with a distinguished set of generators S and a presentation

$$W = \langle s \in S \mid (st)^{m_{st}} = \text{id} \rangle$$

such that $m_{ss} = 1$ and $m_{st} = m_{ts} \in \{2, 3, 4, \dots, \infty\}$ for all $s \neq t$. (We interpret $(st)^\infty = \text{id}$ as there being no Relation.) As we discussed above, the Weyl groups of compact Lie groups are Coxeter groups. In the 1930s Coxeter proved that the finite reflection groups are exactly the finite Coxeter groups, and achieved in this way a classification. As well as the finite reflection groups arising in Lie theory (of types $A, \dots, G$) one has the symmetries of the regular n-gon (a dihedral group of type $I_2(n)$) for $n \neq 3, 4, 6$, the symmetries of the icosahedron (a group of type H_3) and the symmetries of a regular polytope in $\mathbb{R}^4$ with 600 sides (a group of type H_4).

It was realized later (by Coxeter, Tits, ...) that Coxeter groups form an interesting class of groups whether or not they are finite. They encompass groups generated by affine reflections in Euclidean space (affine Weyl groups), certain hyperbolic reflection groups, etc. One can treat these groups in a uniform way thanks to the existence of their *geometric representation*. Let $\mathfrak{h} = \bigoplus_{s \in S} \mathbb{R}\alpha_s^\vee$ for formal symbols $\alpha_s^\vee$, and define a form on $\mathfrak{h}$ via

$$(\alpha_s^\vee, \alpha_t^\vee) = -\cos(\pi/m_{st}).$$

Although this form is positive definite if and only if W is finite, one can still imagine that each $\alpha_s^\vee$ has length 1 and the angle between $\alpha_s^\vee$ and $\alpha_t^\vee$ for $s \neq t$ is $(m_{st} - 1)\pi/m_{st}$. It is not difficult to verify (see [Bou68, V.4.1] or [Hum90, 5.3]) that the assignment

$$s(v) := v - 2(v, \alpha_s^\vee)\alpha_s^\vee$$

defines a representation of W on $\mathfrak{h}$. In fact it is faithful ([Bou68, V.4.4.2] or [Hum90, Corollary 5.4]).

If W happens to be the Weyl group of our $T \subset G$ from the Introduction, then (by rescaling the coroots so that they all have length 1 with respect to a W-invariant form) one may construct a W-equivariant isomorphism

$$\operatorname{Lie} T \cong \mathfrak{h}.$$

Hence one can think of this setup as providing the action of W on the Lie algebra of a maximal torus, even though the corresponding Lie group might not exist!

The main point of the previous section is that one may describe the intersection cohomology, Poincaré pairing and ample cone entirely algebraically, using only $\mathfrak{h}$, its basis and its W-action. That is, let us (re)define $R = S(\mathfrak{h}^*)$ to be the symmetric algebra on $\mathfrak{h}^*$ (alias the regular functions on $\mathfrak{h}$), graded with $\deg \mathfrak{h}^* = 2$. Then W acts on R via graded algebra automorphisms. Imitating the constructions of the previous section one obtains graded R-modules D_w (well-defined up to isomorphism), the only difference being that we work in the category of R-modules rather than $R/(R_+^W)$-modules.[2] We call the modules D_w the (indecomposable) *Soergel modules*. As in the Weyl group case, the modules D_w are finite dimensional over $\mathbb{R}$ and are equipped with non-degenerate "Poincaré pairings":

$$\langle -, - \rangle : D_w^i \times D_w^{2\ell(w)-i} \to \mathbb{R}.$$

Our main theorem is that these modules D_w "look like the intersection cohomology of a Schubert variety". Consider the "ample cone":

$$\mathfrak{h}_+^* := \{\lambda \in \mathfrak{h}^* \mid \langle \lambda, \alpha_s^\vee \rangle > 0 \text{ for all } s \in S\}.$$

Theorem 5.1 ([EW12]). *For any $w \in W$, let D_w be as above.*

(1) *(Hard Lefschetz theorem) For any $i \leq \ell(w)$, left multiplication by λ^i for any $\lambda \in \mathfrak{h}_+^*$ gives an isomorphism*

$$\lambda^{\ell(w)-i} : D_w^i \xrightarrow{\sim} D_w^{2\ell(w)-i}$$

(2) *(Hodge-Riemann bilinear relations) For any $i \leq \ell(w)$ and $\lambda \in \mathfrak{h}_+^*$ the restriction of the form*

$$(f, g) := \langle f, \lambda^{\ell(w)-i} g \rangle$$

on D_w^i to $P^i = \ker \lambda^{\ell(w)-i+1} \subset D_w^i$ is $(-1)^{i/2}$-definite.

Some remarks:

[2] Although all the R-modules will factor through $R/(R_+^W)$, we prefer the ring R for philosophical reasons. When W is infinite, the ring $R/(R_+^W)$ is infinite dimensional, as R^W has the "wrong" transcendence degree, and the Chevalley theorem does not hold. The ring R behaves in a uniform way for all Coxeter groups, while the quotient ring $R/(R_+^W)$ does not.

(1) The graded modules D_w are zero in odd-degree (as is immediate from their definition as a summand of $R \otimes_{R^{s_1}} \cdots \otimes_{R^{s_m}} \mathbb{R}$) and so the sign $(-1)^{i/2}$ makes sense.
(2) The motivation behind establishing the above theorem is a conjecture made by Soergel in [Soe07, Vermutung 1.13]. In fact, the above theorem forms part of a complicated inductive proof of Soergel's conjecture. Soergel was led to his conjecture as an algebraic means of understanding the Kazhdan–Lusztig basis of the Hecke algebra and the Kazhdan–Lusztig conjecture on characters of simple highest weight modules over complex semi-simple Lie algebras. The definition of the Kazhdan–Lusztig basis and the statement of the Kazhdan–Lusztig conjecture is "elementary" but, prior to the above results, needed powerful tools from algebraic geometry (e.g. Deligne's proof of the Weil conjectures) for its resolution. Because of this reliance on algebraic geometry, these methods break down for arbitrary Coxeter systems, for which no flag variety exists. In some sense the above theorem is interesting because it provides a "geometry" for Kazhdan–Lusztig theory for Coxeter groups which do not come from Lie groups or generalizations (affine, Kac–Moody, ...) thereof. This was Soergel's aim in formulating his conjecture.
(3) Our proof is inspired by the beautiful work of de Cataldo and Migliorini [dCM02, dCM05], which proves the decomposition theorem using only classical Hodge theory.
(4) The idea of considering the "intersection cohomology" of a Schubert variety associated with any element in a Coxeter group has also been pursued by Dyer [Dye95, Dye09] and Fiebig [Fie08]. There is also a closely related theory non-rational polytopes (where the associated toric variety is missing) [BL03, Kar04, BF07].
(5) In Example 4.5 we saw that if W is a Weyl group, then an important example of a Soergel module is

$$D_{w_0} \cong R/(R_+^W).$$

In fact this isomorphism holds for any finite Coxeter group W with longest element w_0. The "coinvariant"[3] algebra $R/(R_+^W)$ has been studied by many authors from many points of view. However even in this basic example it seems to be difficult to check the hard Lefschetz theorem or Hodge-Riemann bilinear relations directly. In the next section we will do this by hand when W is a dihedral group.
(6) In [EW12] we work with $\mathfrak{h}$ a slightly larger representation containing the geometric representation. We do this for technical reasons (to ensure that the category of Soergel bimodules is well-behaved). However, one can deduce Theorem 5.1 from the results of [EW12]. The idea of using the results for the

[3] W. Soergel pointed out that this is a bad name, as it has nothing whatsoever to do with coinvariants.

slightly larger representation to deduce results for the geometric representation goes back to Libedinsky [Lib08].

(7) (For the experts.) In [EW12] we prove the results above for certain R-modules $\overline{B_w}$, whose definition differs subtly from that of D_w. However, given that $\overline{B_w}$ is indecomposable as an R-module, one can show easily that $\overline{B_w}$ and D_w are isomorphic. This will be explained elsewhere.

6 The Flag Variety of a Dihedral Group

In this final section we amuse ourselves with the coinvariant ring of a finite dihedral group. We check the hard Lefschetz property and Hodge-Riemann bilinear relations directly.

6.1 *Gauß's q-Numbers*

We start by recalling Gauß's q-numbers. By definition

$$[n] := q^{-n+1} + q^{-n+3} + \cdots + q^{n-3} + q^{n-1} = \frac{q^n - q^{-n}}{q - q^{-1}} \in \mathbb{Z}[q^{\pm 1}].$$

Many identities between numbers can be lifted to identities between q-numbers. We will need

$$[2][n] = [n+1] + [n-1] \tag{4}$$

$$[n]^2 = [2n-1] + [2n-3] + \cdots + [1]. \tag{5}$$

$$[n][n+1] = [2n] + [2n-2] + \cdots + [2]. \tag{6}$$

For the representation theorist, $[n]$ is the character of the simple $\mathfrak{sl}_2(\mathbb{C})$-module of dimension n and the relations above are instances of the Clebsch–Gordan formula.

If $\zeta = e^{2\pi i/2m} \in \mathbb{C}$, then we can specialize $q = \zeta$ to obtain algebraic integers $[n]_\zeta \in \mathbb{R}$. Because $\zeta^m = -1$ we have

$$[m]_\zeta = 0, \quad [i]_\zeta = [m-i]_\zeta, \quad [i+m]_\zeta = -[i]_\zeta. \tag{7}$$

Because ζ^n has positive imaginary part for $n < m$, it is clear that

$$[n]_\zeta \text{ is positive for } 0 < n < m. \tag{8}$$

We use this positivity in a crucial way below. Had we foolishly chosen ζ to be a primitive $2m$th root of unity with non-maximal real part, (8) would fail.

6.2 The Reflection Representation of a Dihedral Group

Now let W be a finite dihedral group of order $2m$. That is $S = \{s_1, s_2\}$ and

$$W = \langle s_1, s_2 \mid s_1^2 = s_2^2 = (s_1 s_2)^m = \mathrm{id}\rangle.$$

Let $\mathfrak{h} = \mathbb{R}\alpha_1^\vee \oplus \mathbb{R}\alpha_2^\vee$ be the geometric representation of (W, S), as in Sect. 5. Because W is finite the form $(-,-)$ on $\mathfrak{h}$ is non-degenerate. We define simple roots $\alpha_1, \alpha_2 \in \mathfrak{h}^*$ by $\alpha_1 = 2(\alpha_1^\vee, -)$ and $\alpha_2 = 2(\alpha_2^\vee, -)$. Then the "Cartan" matrix is

$$(\langle \alpha_i^\vee, \alpha_j\rangle)_{i,j\in\{1,2\}} = \begin{pmatrix} 2 & -\varphi \\ -\varphi & 2 \end{pmatrix} \tag{9}$$

where $\varphi = 2\cos(\pi/m)$. Note that $\varphi = \zeta + \zeta^{-1}$ where $\zeta = e^{2\pi i/2m} \in \mathbb{C}$. Hence $\varphi = [2]_\zeta$ is the notation of the previous section. In particular it is an algebraic integer.

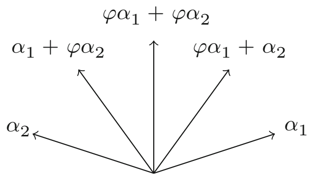
Example 6.1. Throughout we will use the first non-Weyl-group case $m = 5$ to illustrate what is going on. In this case $[2]_\zeta = [3]_\zeta$ and the relation $[2]^2 = [3] + [1]$ gives $\varphi^2 = \varphi + 1$. Thus φ is the golden ratio.

For all $v \in \mathfrak{h}^*$ we have

$$s_1(v) = v - \langle v, \alpha_1^\vee\rangle\alpha_1 \quad \text{and} \quad s_2(v) = v - \langle v, \alpha_2^\vee\rangle\alpha_2.$$

It is a pleasant exercise for the reader to verify that the set $\Phi = W \cdot \{\alpha_1, \alpha_2\}$ gives something like a root system in $\mathfrak{h}^*$. We have $\Phi = \Phi^+ \cup -\Phi^+$ where

$$\Phi^+ = \{[i]_\zeta\alpha_1 + [i-1]_\zeta\alpha_2 \mid 1 \le i \le m\}. \tag{10}$$



Let $T := \bigcup wSw^{-1}$. Then T are precisely the elements of W which act as reflections on $\mathfrak{h}$ (and $\mathfrak{h}^*$). One has a bijection

$$T \xrightarrow{\sim} \Phi^+ : t \mapsto \alpha_t$$

such that $t(\alpha_t) = -\alpha_t$ for all $t \in T$.

6.3 Schubert Calculus

In the following we describe Schubert calculus for the coinvariant ring. Most of what we say here is valid for any finite Coxeter group. A good reference for the unproved statements below is [Hil82].

Let R denote the symmetric algebra on $\mathfrak{h}^*$ and H the coinvariant algebra

$$H := R/(R_+^W).$$

For each $s \in S$ consider the divided difference operator

$$\partial_s(f) = \frac{f - s(f)}{\alpha_s}.$$

Then ∂_s preserves R and decreases degrees by 2. Given $x \in W$ we define

$$\partial_x = \partial_{s_1} \dots \partial_{s_m}$$

where $x = s_1 \dots s_m$ is a reduced expression for x. The operators ∂_s satisfy the braid relations, and therefore ∂_x is well-defined. The operators ∂_x kill invariant polynomials and hence commute with multiplication by invariants. In particular they preserve the ideal (R_+^W) and induce operators on H.

Let $\pi := \prod_{\alpha \in \Phi^+} \alpha$ denote the product of the positive roots. For any $x \in W$ define $Y_x \in H$ as the image of $\partial_x(\pi)$ in H. Because π has degree $2\ell(w_0)$, Y_x has degree $\deg Y_x = 2(\ell(w_0) - \ell(x))$.

Theorem 6.3. *The elements $\{Y_x \mid x \in W\}$ give a basis for H.*

This basis is called the *Schubert basis*. When W is a Weyl group each Y_x maps under the Borel isomorphism to the fundamental class of a Schubert variety [BGG73].

We can define a bilinear form $\langle -, - \rangle$ on H as follows:

$$\langle f, g \rangle := \frac{1}{2m} \partial_{w_0}(fg)$$

Then for all $x, z \in W$ one has:

$$\langle Y_x, Y_z \rangle = \delta_{w_0, x^{-1}z}. \tag{11}$$

In particular $\langle -, - \rangle$ is a non-degenerate form on H.

The following "Chevalley" formula describes the action of an element $f \in \mathfrak{h}^*$ in the basis $\{Y_x\}$:

$$f \cdot Y_x = \sum_{\substack{t \in T \\ \ell(tx) = \ell(x) - 1}} \langle f, \alpha_t^\vee \rangle Y_{tx} \tag{12}$$

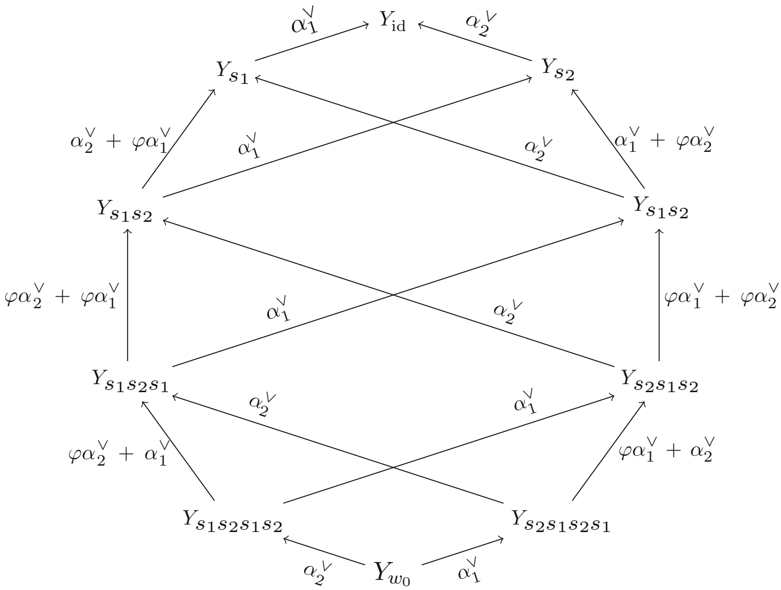


Fig. 1 The Chevalley formula for the dihedral group with $m = 5$

Example 6.4. Figure 1 depicts the case $m = 5$. Each edge is labelled with the coroot which, when paired against f, gives the scalar coefficient that describes the action of f. Using (10) the reader can guess what the picture looks like for general m.

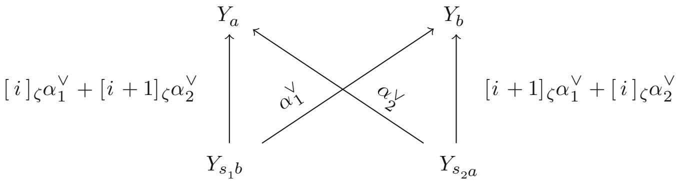
Proposition 6.5. *Suppose that $\lambda \in \mathfrak{h}^*$ is such that $\langle \alpha_i^\vee, \lambda \rangle > 0$ for $i = 1, 2$. Then multiplication by λ on H satisfies the hard Lefschetz theorem, and the Hodge-Riemann bilinear relations hold.*

Proof. It is immediate from (12) that if λ is as in the proposition and if $x \neq \mathrm{id}$ then λY_x is a sum of various Y_z with strictly positive coefficients (two terms occur if $\ell(x) < m - 1$ and one term occurs if $\ell(x) = m - 1$). Hence $\lambda^m Y_{w_0}$ is a strictly positive constant times Y_{id}. In particular $\lambda^m : H^0 = \mathbb{R}Y_{w_0} \to H^{2m} = \mathbb{R}Y_{\mathrm{id}}$ is an isomorphism. By (11) we have

$$\langle Y_{w_0}, \lambda^m Y_{w_0} \rangle > 0$$

and hence the Lefschetz form is positive definite on H^0.

We now fix $1 \leq i < m - 1$ and consider multiplication by $f \in \mathfrak{h}^*$ as a map $H^{2i} \to H^{2i+2}$. The following diagram depicts the effect in the Schubert basis:



(13)

where a and b (resp. s_2a and s_1b) are the unique elements of length $\ell(w_0) - i - 1$ (resp. $\ell(w_0) - i$). Remember that $\alpha_i^\vee$ here represents the scalar $\langle f, \alpha_i^\vee \rangle$. We now calculate the determinant:

$$\det\begin{pmatrix} [i]_\zeta\alpha_1^\vee + [i+1]_\zeta\alpha_2^\vee & \alpha_2^\vee \\ \alpha_1^\vee & [i+1]_\zeta\alpha_1^\vee + [i]_\zeta\alpha_2^\vee \end{pmatrix}$$
$$= [i]_\zeta[i+1]_\zeta(\alpha_1^\vee)^2 + ([i]_\zeta^2 + [i+1]_\zeta^2 - 1)\alpha_1^\vee\alpha_2^\vee + [i]_\zeta[i+1]_\zeta(\alpha_2^\vee)^2$$
$$= [i]_\zeta[i+1]_\zeta(\alpha_1^\vee)^2 + [2]_\zeta[i]_\zeta[i+1]_\zeta\alpha_1^\vee\alpha_2^\vee + [i]_\zeta[i+1]_\zeta(\alpha_2^\vee)^2$$

(using (4), (5) and (6)). All q-numbers appearing here are positive by (8).

If λ is as in the proposition, then the determinant of multiplication by λ is positive. So λ gives an isomorphism $H^{2i} \xrightarrow{\sim} H^{2i+2}$ for each $1 \le i \le m-2$, and λ^{m-2} gives an isomorphism $H^2 \xrightarrow{\sim} H^{2m-2}$. Therefore the hard Lefschetz theorem holds for λ, with primitive classes occurring only in degrees 0 and 2.

It remains to check the Hodge-Riemann bilinear relations. We have already seen that the Lefschetz form on H^0 is positive definite. We need to know that the restriction of the Lefschetz form on H^2 to $\ker\lambda^{m-1}$ is negative definite. Now $(\lambda Y_{w_0}, \lambda Y_{w_0}) = (Y_{w_0}, Y_{w_0}) > 0$, and if $\gamma \in H^2$ denotes a generator for $\ker\lambda^{m-1}$ then $(\lambda Y_{w_0}, \gamma) = \langle \lambda Y_{w_0}, \lambda^{m-2}\gamma \rangle = \langle Y_{w_0}, \lambda^{m-1}\gamma \rangle = 0$. Hence the Hodge-Riemann relations hold if and only if the signature of the Lefschetz form on H^2 is zero.

From the definition of the Lefschetz form, it is immediate that $\lambda : H^{2i} \to H^{2i+2}$ is an isometry with respect to the Lefschetz forms, so long as $2 \le 2i \le m-2$. Thus when m is even (resp. odd) it is enough to show that the signature of the Lefschetz form is zero on H^m (resp. H^{m-1}).

Suppose m is even. The Lefschetz form on the middle dimension H^m is the same as the pairing. By (11) this form has Gram matrix

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

which has signature 0.

Suppose $m = 2k+1$ is odd; we check the signature of the Lefschetz form on H^{m-1}. We are reduced to studying 13 with $\ell(a) = \ell(b) = k$ and $\ell(s_2a) = \ell(s_1b) = k+1$. We see by (11) that Y_{s_1b}, Y_{s_2a} is a basis dual to Y_b, Y_a. We get that the Lefschetz form on H^{m-1} is given by

$$\begin{pmatrix} \alpha_1^\vee & [k+1]_\zeta\alpha_1^\vee + [k]_\zeta\alpha_2^\vee \\ [k]_\zeta\alpha_1^\vee + [k+1]_\zeta\alpha_2^\vee & \alpha_2^\vee \end{pmatrix},$$

and $[k] = [k+1]$ is positive. For any λ as in the proposition, this is a symmetric matrix with strictly positive entries and negative determinant (by our calculation above). Hence its signature is zero and the Hodge-Riemann relations are satisfied as claimed. □

References

[Ara06] A. Arabia, Correspondance de Springer. Preprint (2006) http://www.institut.math.jussieu.fr/?arabia/math/Pervers.pdf

[BBD82] A. Beĭlinson, J. Bernstein, P. Deligne, Faisceaux pervers, in *Analyse et topologie sur les espaces singuliers, I (Luminy, 1981)*. Astérisque, vol. 100 (Société Mathématique de France, Paris, 1982), pp. 5–171

[BF07] G. Barthel, J.-P. Brasselet, K.-H. Fieseler, L. Kaup, Hodge-Riemann relations for polytopes. A geometric approach, in *Singularity theory. Proceedings of the 2005 Marseille singularity school and conference, CIRM, Marseille, France, January 24–February 25, 2005. Dedicated to Jean-Paul Brasselet on his 60th birthday* (World Scientific, Singapore, 2007), pp. 379–410

[BGG73] I.N. Bernšteĭn, I.M. Gel′fand, S.I. Gel′fand. Schubert cells, and the cohomology of the spaces G/P. Uspehi Mat. Nauk **28**(3(171)), 3–26 (1973)

[BL03] P. Bressler, V.A. Lunts, Intersection cohomology on nonrational polytopes. Compos. Math. **135**(3), 245–278 (2003)

[Bor94] A. Borel, Introduction to middle intersection cohomology and perverse sheaves, in *Algebraic Groups and Their Generalizations: Classical Methods. Summer Research Institute on Algebraic Groups and Their Generalizations, July 6–26, 1991, Pennsylvania State University, University Park, PA, USA* (American Mathematical Society, Providence, 1994), pp. 25–52

[Bou68] N. Bourbaki, *Éléments de mathématique. Fasc. XXXIV. Groupes et algèbres de Lie. Chapitre IV: Groupes de Coxeter et systèmes de Tits. Chapitre V: Groupes engendrés par des réflexions. Chapitre VI: systèmes de racines*. Actualités Scientifiques et Industrielles, No. 1337 (Hermann, Paris, 1968)

[Bri12] M. Brion, Lectures on the geometry of flag varieties (2012). arXiv:0410240

[BS58] R. Bott, H. Samelson, Applications of the theory of Morse to symmetric spaces. Am. J. Math. **80**, 964–1029 (1958)

[dCM02] M.A.A. de Cataldo, L. Migliorini, The hard Lefschetz theorem and the topology of semismall maps. Ann. Sci. École Norm. Sup. (4) **35**(5), 759–772 (2002)

[dCM05] M.A.A. de Cataldo, L. Migliorini, The Hodge theory of algebraic maps. Ann. Sci. École Norm. Sup. (4) **38**(5), 693–750 (2005)

[Dem74] M. Demazure, Désingularisation des variétés de Schubert généralisées. Ann. Sci. École Norm. Sup. (4) **7**, 53–88 (1974). Collection of articles dedicated to Henri Cartan on the occasion of his 70th birthday, I

[dM09] M.A.A. de Cataldo, L. Migliorini, The decomposition theorem, perverse sheaves and the topology of algebraic maps. Bull. Am. Math. Soc., New Ser. **46**(4), 535–633 (2009)

[Dye95] M. Dyer, Representation theories from Coxeter groups, in *Representations of Groups. Canadian Mathematical Society Annual Seminar, June 15–24, 1994, Banff, Alberta, Canada* (American Mathematical Society, Providence, 1995), pp. 105–139

[Dye09] M. Dyer, Modules for the dual nil Hecke ring (2009). Preprint

[EW12] B. Elias, G. Williamson, The Hodge theory of Soergel bimodules. Ann. Math. (2012, to appear). arXiv:1212.0791

[Fie08] P. Fiebig, The combinatorics of Coxeter categories. Trans. Am. Math. Soc. **360**(8), 4211–4233 (2008)

[GM80] M. Goresky, R. MacPherson, Intersection homology theory. Topology **19**, 135–165 (1980)

[GM83] M. Goresky, R. MacPherson, Intersection homology. II. Invent. Math. **72**, 77–129 (1983)

[Han73] H.C. Hansen, On cycles in flag manifolds. Math. Scand. **33**, 269–274 (1974) (1973)

[Hil82] H. Hiller, Geometry of Coxeter groups. *Research Notes in Mathematics*, vol. 54 (Pitman Advanced Publishing Program, Boston, 1982), p. 213

[Hum90] J.E. Humphreys, Reflection groups and Coxeter groups. *Cambridge Studies in Advanced Mathematics*, vol. 29 (Cambridge University Press, Cambridge, 1990)

[Kar04] K. Karu, Hard Lefschetz theorem for nonrational polytopes. Invent. Math. **157**(2), 419–447 (2004)

[Kle07] S.L. Kleiman, The development of intersection homology theory. Pure Appl. Math. Q. **3**(1), 225–282 (2007)

[Lib08] N. Libedinsky, Équivalences entre conjectures de Soergel. J. Algebra **320**(7), 2695–2705 (2008)

[Rie04] K. Rietsch, An introduction to perverse sheaves, in *Representations of Finite Dimensional Algebras and Related Topics in Lie Theory and Geometry. Proceedings from the 10th International Conference on Algebras and Related Topics, ICRA X, Toronto, Canada, July 15–August 10, 2002* (American Mathematical Society (AMS), Providence, 2004), pp. 391–429

[Sai89] M. Saito, Introduction to mixed Hodge modules. *Théorie de Hodge, Actes Colloq., Luminy/Fr. 1987*. Astérisque 179–180 (1989), pp. 145–162

[Soe90] W. Soergel. Kategorie $\mathcal{O}$, perverse Garben und Moduln über den Koinvarianten zur Weylgruppe. J. Am. Math. Soc., **3**(2), 421–445 (1990)

[Soe07] W. Soergel, Kazhdan-Lusztig-Polynome und unzerlegbare Bimoduln über Polynomringen. J. Inst. Math. Jussieu **6**(3), 501–525 (2007)

Uniform Sup-Norm Bounds on Average for Cusp Forms of Higher Weights

Joshua S. Friedman, Jay Jorgenson, and Jürg Kramer

Dedicated to the Memory of Friedrich Hirzebruch

Abstract Let $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ be a Fuchsian subgroup of the first kind acting by fractional linear transformations on the upper half-plane $\mathbb{H}$. Consider the d-dimensional space of cusp forms $\mathcal{S}_{2k}^{\Gamma}$ of weight $2k$ for Γ, and let $\{f_1, \ldots, f_d\}$ be an orthonormal basis of $\mathcal{S}_{2k}^{\Gamma}$ with respect to the Petersson inner product. In this paper we show that the sup-norm of the quantity $S_{2k}^{\Gamma}(z) := \sum_{j=1}^{d} |f_j(z)|^2 \operatorname{Im}(z)^{2k}$ is bounded as $O_{\Gamma}(k)$ in the cocompact setting, and as $O_{\Gamma}(k^{3/2})$ in the cofinite case, where the implied constants depend solely on Γ. We also show that the implied constants are uniform if Γ is replaced by a subgroup of finite index.

The views expressed in this article are the author's own and not those of the U.S. Merchant Marine Academy, the Maritime Administration, the Department of Transportation, or the United States government.

J.S. Friedman
Department of Mathematics and Science, United States Merchant Marine Academy,
300 Steamboat Road, Kings Point, NY 11024, USA
e-mail: FriedmanJ@usmma.edu

J. Jorgenson
Department of Mathematics, The City College of New York, Convent Avenue at 138th Street,
New York, NY 10031, USA
e-mail: jjorgenson@mindspring.com

J. Kramer (✉)
Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6,
10099 Berlin, Germany
e-mail: kramer@math.hu-berlin.de

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_6

1 Introduction

1.1 Motivation

Let M denote a compact Riemann surface of genus $g \geq 2$. From the uniformization theorem there is a unique metric on M, which is compatible with its complex structure and which has constant Gauss curvature equal to -1. On the other hand, from complex algebraic geometry, there is a canonical metric on M obtained by pull-back through the Abel–Jacobi map from M into its Jacobian variety $\mathrm{Jac}(M)$. Let μ_{hyp} and μ_{can} denote the $(1,1)$-forms associated with the hyperbolic and canonical metrics, respectively. A natural question to consider is to compare μ_{hyp} with μ_{can}, in whatever manner possible for general compact Riemann surfaces M of genus $g \geq 2$. Since M has volume 1 with respect to μ_{can}, let us rescale the hyperbolic metric by a multiplicative constant so that the associated $(1,1)$-form μ_{shyp} also gives M volume 1. After some reflection upon the question in hand, one concludes that perhaps the most approachable manner in which one can compare the two metrics is to consider the sup-norm of the function $\mu_{\mathrm{can}}/\mu_{\mathrm{shyp}}$ on M viewed as a finite degree cover of some fixed base Riemann surface M_0.

In [JK04], the authors proved the following (optimal) result. If M is a compact Riemann surface of genus $g \geq 2$, which is a finite degree cover of a fixed compact Riemann surface M_0, then the bound

$$\sup_{z \in M} \left(\frac{\mu_{\mathrm{can}}(z)}{\mu_{\mathrm{shyp}}(z)} \right) = O_{M_0}(1) \tag{1}$$

holds. To be precise, the main result of [JK04] applies whenever M is a finite degree cover of M_0, which has finite hyperbolic volume and need not necessarily be compact. In the setting of arithmetic geometry, the ratio $\sup_{z \in M} \mu_{\mathrm{can}}(z)/\mu_{\mathrm{shyp}}(z)$ appears as an analytic invariant in the Arakelov theory of algebraic curves, and the bound (1) plays an important role in [JK09], where the authors derived bounds for Faltings's delta function and, subsequently, for the Faltings height of Jacobians associated with modular curves. Further comments on the significance of (1) as well as related results will be given in Sect. 1.3 below.

From the point of view of automorphic forms, the ratio $\mu_{\mathrm{can}}/\mu_{\mathrm{shyp}}$ roughly equals the sum of squared norms of an orthonormal basis of cusp forms of weight 2 on M. From this point of view, we can extend the bound (1) in two regards: first, we can consider the sum of squared norms of an orthonormal basis of cusp forms of arbitrary weight $2k$ on M, and second, we can develop bounds which are uniform in the weight. The study of these two questions is the subject of the present article.

1.2 Statement of Results

Let $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ be a Fuchsian subgroup of the first kind acting by fractional linear transformations on the upper half-plane $\mathbb{H}$, and let $M := \Gamma\backslash\mathbb{H}$ be the corresponding quotient space. We then consider the $\mathbb{C}$-vector space $\mathcal{S}_{2k}^{\Gamma}$ of cusp forms of weight $2k$ for Γ, and let $\{f_1, \ldots, f_d\}$ be an orthonormal basis of $\mathcal{S}_{2k}^{\Gamma}$ with respect to the Petersson inner product; here $d := \dim_{\mathbb{C}}(\mathcal{S}_{2k}^{\Gamma})$. With these notations, we put for $z \in \mathbb{H}$

$$S_{2k}^{\Gamma}(z) := \sum_{j=1}^{d} |f_j(z)|^2 \operatorname{Im}(z)^{2k}.$$

In this article, we prove optimal L^{∞}-bounds for $S_{2k}^{\Gamma}(z)$ in two different directions, namely uniform L^{∞}-bounds with regard to the weight $2k$, as well as uniform L^{∞}-bounds through finite degree covers of M. More precisely, the following statement is proven:

Let $\Gamma_0 \subset \mathrm{PSL}_2(\mathbb{R})$ be a fixed Fuchsian subgroup of the first kind and let $\Gamma \subseteq \Gamma_0$ be any subgroup of finite index. For any $k \in \mathbb{N}_{>0}$, we then have the bound

$$\sup_{z\in M}\left(S_{2k}^{\Gamma}(z)\right) = O_{\Gamma_0}(k^{3/2}), \tag{2}$$

where the implied constant depends solely on Γ_0. Moreover, if Γ_0 is cocompact, then we have the improved bound

$$\sup_{z\in M}\left(S_{2k}^{\Gamma}(z)\right) = O_{\Gamma_0}(k), \tag{3}$$

where, again, the implied constant depends solely on Γ_0.

We were somewhat surprised to find different orders of growth in the weight comparing the cocompact to the general cofinite case. With regard to this phenomenon, we prove the following auxiliary result in Proposition 5.1, which indicates where the maximal values occur in the cofinite case:

For a cofinite Fuchsian subgroup Γ of the first kind and $k \in \mathbb{N}_{>0}$, let $\varepsilon > 0$ be such that the neighborhoods of area ε around the cusps of M are disjoint. Assuming that $0 < \varepsilon < 2\pi/k$, we have the bound

$$\sup_{z\in M}\left(S_{2k}^{\Gamma}(z)\right) = O_{\Gamma,\varepsilon}(k),$$

where the implied constant depends solely on Γ and ε.

Moreover, as far as the bounds (2) and (3) are concerned, we are able to show that the results are optimal in both cases, at least up to an additive term in the exponent of the form $-\varepsilon$ for any $\varepsilon > 0$.

1.3 Related Results

As stated, the origin of the problems considered in the present article comes from [JK04], which studies the case of cusp forms of weight 2, i.e., $k = 1$ in the present notation. In this respect, we recall that in the case $\Gamma_0 = \mathrm{PSL}_2(\mathbb{Z})$ and $\Gamma = \Gamma_0(N)$, as a first step the main result of [AU95] proved for any $\varepsilon > 0$ that

$$\sup_{z\in M}\left(S_2^{\Gamma_0(N)}(z)\right) = O(N^{2+\varepsilon}),$$

which was improved to $O(N^{1+\varepsilon})$ in [MU98]. In [JK04], the bound was finally improved to $O_{\Gamma_0}(1)$, not only for the above-mentioned setting, but also to the case when neither Γ nor Γ_0 possess any arithmetic properties. With this stated, the present article generalizes the results of [JK04] to cusp forms of arbitrary even weight and for arbitrary Fuchsian subgroups $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ of the first kind.

In a related direction, there has been considerable interest in obtaining sup-norm bounds for individual Hecke eigenforms, with the most recent results coming from the setting when the groups under consideration are arithmetic. For example, the holomorphic setting of the quantum unique ergodicity (QUE) problem has been studied in [LS03, Lau] and [HS10]. In [HS10], it is proven for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ that normalized Hecke eigenforms of weight $2k$ converge weakly to the constant function $3/\pi$ as k tends to infinity. In another direction, the authors prove in [HT13] the so far best known bound for the L^∞-norm of L^2-normalized Hecke eigenforms for the congruence subgroups $\Gamma_0(N)$ for squarefree N. Specifically, it is shown that

$$\|f\|_\infty \ll_\varepsilon k^{\frac{11}{2}} N^{-\frac{1}{6}+\varepsilon},$$

with an implied constant which only depends on $\varepsilon > 0$. We refer to the introduction as well as the bibliography of the paper [HT13], which gives an excellent account on the improvements of the bounds for the L^∞-norm of L^2-normalized Hecke eigenforms for the congruence subgroups $\Gamma_0(N)$.

When comparing the results of the above articles to the main theorem of [JK04] and the present article, one comes to the conclusion that the various results are complementary. From the main result in the present paper in the case $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$, one obtains a bound for individual cusp forms which is weaker than in the theorems of the above-mentioned articles. When taking the average results from the above-mentioned articles, one obtains an average bound which is weaker than the main theorem in the present paper.

More recently, a number of articles have appeared whose results are closely related to the contents of [JK04, JK13], or the present article. In [Javanpeykar, Kaenel1, Kaenel2], and [Pazuki], the authors study various fundamental problems in arithmetic, such as Shafarevich-type conjectures using certain aspects of Arakelov theory, including bounds for certain analytic invariants such as (1) as well as effective bounds for Faltings's delta function (see [JK13]). In [BF13], the authors prove an arithmetic analogue of the Hilbert–Samuel theorem, which has far-reaching potential; the main result of the present article is related to the Bergman measure studied in Sect. 2.5 therein.

Finally, we refer the reader to the interesting article [Templier], in which the author proves the existence of cusp forms which, in the (not necessarily squarefree) level aspect, have large modulus, thus disproving a "folklore" conjecture asserting that all forms should be uniformly small.

1.4 Outline of the Paper

In Sect. 2, we establish notations and recall background material. In Sect. 3, we prove technical results for the heat kernel associated with the Laplacian Δ_k acting on Maass forms of weight k for Γ. In Sect. 4, we provide a proof of the bound (3) for $\Gamma = \Gamma_0$. By an additional investigation in the neighborhoods of the cusps, we arrive in Sect. 5 at a proof of the bound (2), again in the case that $\Gamma = \Gamma_0$. Finally, in Sect. 6, we are able to establish the uniformity of our bounds (2) and (3) with regard to finite index subgroups Γ in Γ_0. To complete the article, we show that our bounds are optimal, which is the content of Sect. 7.

2 Background Material

2.1 Hyperbolic Metric

Let $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ be any Fuchsian subgroup of the first kind acting by fractional linear transformations on the upper half-plane $\mathbb{H} := \{z \in \mathbb{C} \mid z = x+iy\,,\ y > 0\}$. Let M be the quotient space $\Gamma\backslash\mathbb{H}$ and g the genus of M. Denote by $\mathcal{T}$ the set of elliptic fixed points of M and by $\mathcal{C}$ the set of cusps of M; we put $t := |\mathcal{T}|$ and $c := |\mathcal{C}|$. If $p \in \mathcal{T}$, we let m_p denote the order of the elliptic fixed point p; we set $m_p = 1$, if p is a regular point of M. Locally, away from the elliptic fixed points, we identify M with its universal cover $\mathbb{H}$, and hence, denote the points on $M \setminus \mathcal{T}$ by the same letter as the points on $\mathbb{H}$.

We denote by $\mathrm{d}s^2_{\mathrm{hyp}}(z)$ the line element and by $\mu_{\mathrm{hyp}}(z)$ the volume form corresponding to the hyperbolic metric on M, which is compatible with the complex structure of M and has constant curvature equal to -1. Locally on $M \setminus \mathcal{T}$, we have

$$\mathrm{d}s^2_{\mathrm{hyp}}(z) = \frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2} \quad \text{and} \quad \mu_{\mathrm{hyp}}(z) = \frac{\mathrm{d}x \wedge \mathrm{d}y}{y^2}\,.$$

We denote the hyperbolic distance between $z, w \in M$ by $\mathrm{dist}_{\mathrm{hyp}}(z, w)$ and we recall that the hyperbolic volume $\mathrm{vol}_{\mathrm{hyp}}(M)$ of M is given by the formula

$$\mathrm{vol}_{\mathrm{hyp}}(M) = 2\pi\Bigg(2g - 2 + c + \sum_{p\in\mathcal{T}}\Big(1 - \frac{1}{m_p}\Big)\Bigg).$$

2.2 Cusp Forms of Higher Weights

For $k \in \mathbb{N}_{>0}$, we let $\mathcal{S}_{2k}^{\Gamma}$ denote the space of cusp forms of weight $2k$ for Γ, i.e., the space of holomorphic functions $f : \mathbb{H} \longrightarrow \mathbb{C}$, which have the transformation behavior

$$f(\gamma z) = (cz + d)^{2k} f(z)$$

for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, and which vanish at all the cusps of M. The space $\mathcal{S}_{2k}^{\Gamma}$ is equipped with the inner product

$$\langle f_1, f_2\rangle := \int\limits_M f_1(z)\overline{f_2(z)}\, y^{2k} \mu_{\mathrm{hyp}}(z) \qquad (f_1, f_2 \in \mathcal{S}_{2k}^{\Gamma}).$$

By letting $d := \dim_{\mathbb{C}}(\mathcal{S}_{2k}^{\Gamma})$ and choosing an orthonormal basis $\{f_1, \ldots, f_d\}$ of $\mathcal{S}_{2k}^{\Gamma}$, we define the quantity

$$S_{2k}^{\Gamma}(z) := \sum_{j=1}^{d} |f_j(z)|^2\, y^{2k}.$$

The main result of this paper consists in giving optimal bounds for the quantity $S_{2k}^{\Gamma}(z)$ as z ranges throughout M.

2.3 Maass Forms of Higher Weights

Following [Fay] or [Fischer], we introduce for any $k \in \mathbb{N}$ the space $\mathcal{V}_k^{\Gamma}$ of functions $\varphi : \mathbb{H} \longrightarrow \mathbb{C}$, which have the transformation behavior

$$\varphi(\gamma z) = \left(\frac{cz + d}{c\bar{z} + d}\right)^k \varphi(z) = e^{2ik \arg\,(cz+d)}\varphi(z)$$

for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. For $\varphi \in \mathcal{V}_k^\Gamma$, we set

$$\|\varphi\|^2 := \int_M |\varphi(z)|^2 \mu_{\mathrm{hyp}}(z),$$

whenever it is defined. We then introduce the Hilbert space

$$\mathcal{H}_k^\Gamma := \left\{\varphi \in \mathcal{V}_k^\Gamma \;\middle|\; \|\varphi\| < \infty\right\}$$

equipped with the inner product

$$\langle \varphi_1, \varphi_2 \rangle := \int_M \varphi_1(z)\overline{\varphi_2(z)}\mu_{\mathrm{hyp}}(z) \qquad (\varphi_1, \varphi_2 \in \mathcal{H}_k^\Gamma).$$

The generalized Laplacian

$$\Delta_k := -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + 2iky\frac{\partial}{\partial x}$$

acts on the smooth functions of $\mathcal{H}_k^\Gamma$ and extends to an essentially self-adjoint linear operator acting on a dense subspace of $\mathcal{H}_k^\Gamma$.

From [Fay] or [Fischer], we quote that the eigenvalues for the equation

$$\Delta_k \varphi(z) = \lambda\varphi(z) \qquad (\varphi \in \mathcal{H}_k^\Gamma)$$

satisfy the inequality $\lambda \geq k(1-k)$.

Furthermore, if $\lambda = k(1-k)$, then the corresponding eigenfunction φ is of the form $\varphi(z) = f(z)y^k$, where f is a cusp form of weight $2k$ for Γ, i.e., we have an isomorphism of $\mathbb{C}$-vector spaces

$$\ker\left(\Delta_k - k(1-k)\right) \cong \mathcal{S}_{2k}^\Gamma.$$

2.4 Heat Kernels of Higher Weights

The heat kernel on $\mathbb{H}$ associated with Δ_k is computed in [Oshima] and corrects a corresponding formula in [Fay]. It is given by

$$K_k(t;\rho) = \frac{\sqrt{2}\,e^{-t/4}}{(4\pi t)^{3/2}} \int_\rho^\infty \frac{re^{-r^2/(4t)}}{\sqrt{\cosh(r) - \cosh(\rho)}}\, T_{2k}\left(\frac{\cosh(r/2)}{\cosh(\rho/2)}\right)\mathrm{d}r\,,$$

where

$$T_{2k}(X) := \cosh(2k\,\mathrm{arccosh}(X))$$

denotes the $2k$th Chebyshev polynomial.

The heat kernel on M associated with Δ_k is defined by (see [Fay, p. 153])

$$K_k^{\Gamma}(t; z, w) := \sum_{\gamma \in \Gamma} \left(\frac{c\bar{w} + d}{cw + d} \right)^k \left(\frac{z - \gamma \bar{w}}{\gamma w - \bar{z}} \right)^k K_k(t; \rho_{\gamma; z,w}),$$

where $\rho_{\gamma;z,w} := \mathrm{dist}_{\mathrm{hyp}}(z, \gamma w)$. If $z = w$, we put $\rho_{\gamma;z} := \rho_{\gamma;z,z}$ and $K_k^{\Gamma}(t; z) := K_k^{\Gamma}(t; z, z)$.

2.5 Spectral Expansions

The resolvent kernel on M associated with Δ_k is the integral kernel $G_k^{\Gamma}(s; z, w)$, which inverts the operator $\Delta_k - s(1-s)$ (see [Fischer, p. 27, Theorem 1.4.10]). The heat kernel and the resolvent kernel on M associated with Δ_k are related through the expression

$$G_k^{\Gamma}(s; z, w) = \int_0^{\infty} e^{-(s-1/2)^2 t} e^{t/4} K_k^{\Gamma}(t; z, w), \tag{4}$$

which holds for $s \in \mathbb{C}$ such that $\mathrm{Re}((s - 1/2)^2)$ is sufficiently large. In other words, (4) expresses the resolvent kernel on M associated with Δ_k as the Laplace transform of the heat kernel on M associated with Δ_k, with an appropriate change of variables. Conversely, one then can express the heat kernel on M as an inverse Laplace transform, with an appropriate change of variables, of the resolvent kernel on M.

The spectral expansion of the resolvent kernel on M associated with Δ_k is given on p. 40 of [Fischer], which is established as an example of a more general spectral expansion theorem given on p. 37 of [Fischer]. Using the inverse Laplace transform, one then obtains the spectral expansion for the heat kernel on M associated with Δ_k; we leave the details for the derivation to the interested reader. For the purposes of the present article, we derive from the spectral expansion of $K_k^{\Gamma}(t; z)$ and the fact that the smallest eigenvalue of Δ_k is given by $k(1-k)$ and that the corresponding eigenfunctions are related to $\mathcal{S}_{2k}^{\Gamma}$, the important relation

$$S_{2k}^{\Gamma}(z) = \lim_{t \to \infty} e^{-k(k-1)t} K_k^{\Gamma}(t; z).$$

Furthermore, it is evident from the spectral expansion of the heat kernel that $e^{-k(k-1)t} K_k^{\Gamma}(t; z)$ is a monotone decreasing function for any $t > 0$, hence we arrive at the estimate

$$e^{k(k-1)t} S_{2k}^{\Gamma}(z) \leq K_k^{\Gamma}(t; z) \tag{5}$$

for any $t > 0$ and $z \in \mathbb{H}$.

3 Heat Kernel Analysis

Lemma 3.1. *For $t > 0$, $\rho > 0$, and $r \geq \rho$, let*

$$F_k(t;\rho,r) := \frac{re^{-r^2/(4t)}}{\sinh(r)} T_{2k}\left(\frac{\cosh(r/2)}{\cosh(\rho/2)}\right).$$

Then, for all values of t, ρ, r in the given range, we have

$$\sinh(r)\frac{\partial}{\partial\rho}F_k(t;\rho,r) + \sinh(\rho)\frac{\partial}{\partial r}F_k(t;\rho,r) < 0.$$

Proof. We put

$$X := \frac{\cosh(r/2)}{\cosh(\rho/2)},$$

and compute

$$\begin{aligned}
&\sinh(r)\frac{\partial}{\partial\rho}F_k(t;\rho,r) + \sinh(\rho)\frac{\partial}{\partial r}F_k(t;\rho,r) \\
&= \sinh(\rho)\, F_k(t;\rho,r)\left(\frac{1}{r} - \frac{r}{2t} - \frac{\cosh(r)}{\sinh(r)}\right) \\
&\quad + \frac{re^{-r^2/(4t)}}{\sinh(r)} T'_{2k}(X)\left(\sinh(r)\frac{\partial X}{\partial\rho} + \sinh(\rho)\frac{\partial X}{\partial r}\right).
\end{aligned}$$

It is now easy to see that

$$\frac{1}{r} - \frac{r}{2t} - \frac{\cosh(r)}{\sinh(r)} < 0$$

for all $t > 0$ and $r > 0$. Since $r \geq \rho$, we have $X \geq 1$, and hence

$$T_{2k}(X) = \cosh(2k\,\mathrm{arccosh}(X)) \geq 1,$$

from which we conclude that

$$\sinh(\rho)\, F_k(t;\rho,r)\left(\frac{1}{r} - \frac{r}{2t} - \frac{\cosh(r)}{\sinh(r)}\right) < 0.$$

Furthermore, since $T_{2k}(X)$ is an increasing, positive function, its derivative $T'_{2k}(X)$ is again a positive function. To complete the proof of the lemma, we are therefore left to show that

$$\sinh(r)\frac{\partial X}{\partial \rho} + \sinh(\rho)\frac{\partial X}{\partial r} \leq 0.$$

For this we compute

$$\begin{aligned}
&\sinh(r)\frac{\partial X}{\partial \rho} + \sinh(\rho)\frac{\partial X}{\partial r}\\
&\quad = -\sinh(r)\frac{\cosh(r/2)\sinh(\rho/2)}{2\cosh^2(\rho/2)} + \sinh(\rho)\frac{\sinh(r/2)}{2\cosh(\rho/2)}\\
&\quad = \frac{1}{2\cosh^2(\rho/2)}\big(-\sinh(r)\cosh(r/2)\sinh(\rho/2) + \sinh(\rho)\cosh(\rho/2)\sinh(r/2)\big)\\
&\quad = \frac{1}{2\cosh^2(\rho/2)}\big(-2\sinh(r/2)\cosh^2(r/2)\sinh(\rho/2)\\
&\qquad + 2\sinh(\rho/2)\cosh^2(\rho/2)\sinh(r/2)\big)\\
&\quad = \frac{\sinh(r/2)\sinh(\rho/2)}{\cosh^2(\rho/2)}\big(-\cosh^2(r/2) + \cosh^2(\rho/2)\big),
\end{aligned}$$

which is negative for $r > \rho$ and vanishes for $r = \rho$. □

Proposition 3.2. *For any $t > 0$, the heat kernel $K_k(t;\rho)$ on $\mathbb{H}$ associated with Δ_k is strictly monotone decreasing for $\rho > 0$.*

Proof. We will prove that $\partial/\partial\rho\, K_k(t;\rho) < 0$ for $\rho > 0$. To simplify notations, we put

$$c(t) := \frac{\sqrt{2}e^{-t/4}}{(4\pi t)^{3/2}}.$$

In the notation of Lemma 3.1, we then have, using integration by parts,

$$\begin{aligned}
K_k(t;\rho) &= c(t)\int_\rho^\infty F_k(t;\rho,r)\frac{\sinh(r)}{\sqrt{\cosh(r)-\cosh(\rho)}}\,\mathrm{d}r\\
&= -2c(t)\int_\rho^\infty \frac{\partial}{\partial r}F_k(t;\rho,r)\sqrt{\cosh(r)-\cosh(\rho)}\,\mathrm{d}r\,.
\end{aligned}$$

We now apply the Leibniz rule of differentiation to write

$$\begin{aligned}
\frac{\partial}{\partial\rho}K_k(t;\rho) &= -2c(t)\int_\rho^\infty \frac{\partial^2}{\partial r\,\partial\rho}F_k(t;\rho,r)\sqrt{\cosh(r)-\cosh(\rho)}\,\mathrm{d}r\\
&\quad + c(t)\int_\rho^\infty \frac{\partial}{\partial r}F_k(t;\rho,r)\frac{\sinh(\rho)}{\sqrt{\cosh(r)-\cosh(\rho)}}\,\mathrm{d}r\,.
\end{aligned}$$

Using integration by parts on the first term once again, yields the identity

$$\frac{\partial}{\partial\rho}K_k(t;\rho) = c(t)\int_{\rho}^{\infty}\left(\sinh(r)\frac{\partial}{\partial\rho}F_k(t;\rho,r) + \sinh(\rho)\frac{\partial}{\partial r}F_k(t;\rho,r)\right) \times \frac{\mathrm{d}r}{\sqrt{\cosh(r)-\cosh(\rho)}}.$$

With Lemma 3.1 we conclude that $\partial/\partial\rho\, K_k(t;\rho) < 0$ for $\rho > 0$, which proves the claim. □

Proposition 3.3. *For given Γ, $k \in \mathbb{N}$, and $t > 0$, the heat kernel $K_k^{\Gamma}(t;z)$ on M associated with Δ_k converges absolutely and uniformly on compact subsets K of M.*

Proof. Let $K \subseteq M$ be a compact subset. In order to prove the absolute and uniform convergence of the heat kernel $K_k^{\Gamma}(t;z)$ on M associated with Δ_k for $t > 0$ and $z \in K$, we have to show the convergence of

$$\sum_{\gamma\in\Gamma} K_k(t;\rho_{\gamma;z})$$

for $t > 0$ and $z \in K$. To do this, we introduce for $\rho > 0$ and $z \in K$ the counting function

$$N(\rho;z) := \#\big\{\gamma \in \Gamma \mid \rho_{\gamma;z} = \mathrm{dist}_{\mathrm{hyp}}(z,\gamma z) < \rho\big\}. \tag{6}$$

By arguing as in the proof of Lemma 2.3 (a) of [JL95], one proves that

$$N(\rho;z) = O_{\Gamma,K}(e^{\rho}), \tag{7}$$

uniformly for all $z \in K$ with an implied constant depending solely on Γ and K. The dependence on Γ is given by the maximal order of elliptic elements of Γ.

By means of the counting function $N(\rho;z)$, we obtain the following Stieltjes integral representation of the quantity under consideration:

$$\sum_{\gamma\in\Gamma} K_k(t;\rho_{\gamma;z}) = \int_0^{\infty} K_k(t;\rho)\,\mathrm{d}N(\rho;z).$$

Since $K_k(t;\rho)$ is a non-negative, continuous, and, by Proposition 3.2, monotone decreasing function of ρ, an elementary argument allows one to derive from (7) the bound

$$\int_0^\infty K_k(t;\rho)\,\mathrm{d}N(\rho;z) = O_{\Gamma,K}\left(\int_0^\infty K_k(t;\rho)\,e^\rho\,\mathrm{d}\rho\right), \tag{8}$$

again uniformly for all $z \in K$ with an implied constant depending solely on Γ and K.

We are thus left to find a suitable bound for $K_k(t;\rho)$. For this we observe the inequality

$$\frac{r^2}{4t} \geq \frac{r^2}{8t} + \frac{\rho^2}{8t}$$

for $r \geq \rho$, which gives

$$\begin{aligned} K_k(t;\rho) &= \frac{\sqrt{2}\,e^{-t/4}}{(4\pi t)^{3/2}} \int_\rho^\infty \frac{re^{-r^2/(4t)}}{\sqrt{\cosh(r)-\cosh(\rho)}}\,T_{2k}\left(\frac{\cosh(r/2)}{\cosh(\rho/2)}\right)\mathrm{d}r \\ &\leq e^{-\rho^2/(8t)}\,\frac{\sqrt{2}\,e^{-t/4}}{(4\pi t)^{3/2}} \int_\rho^\infty \frac{re^{-r^2/(8t)}}{\sqrt{\cosh(r)-\cosh(\rho)}}\,T_{2k}\left(\frac{\cosh(r/2)}{\cosh(\rho/2)}\right)\mathrm{d}r \\ &\leq e^{-\rho^2/(8t)}\,\frac{\sqrt{2}\,e^{-t/4}}{(4\pi t)^{3/2}} \int_0^\infty \frac{re^{-r^2/(8t)}}{\sqrt{\cosh(r)-1}}\,T_{2k}(\cosh(r/2))\,\mathrm{d}r\,; \end{aligned} \tag{9}$$

for the last inequality we used that the preceding integral is monotone decreasing in ρ, which follows along the same lines as the proof of Proposition 3.2. Using the equalities

$$\cosh(r) - 1 = 2\sinh^2(r/2) \quad \text{and} \quad T_{2k}(\cosh(r/2)) = \cosh(kr)\,,$$

the estimate (9) leads to the bound

$$K_k(t;\rho) \leq e^{-\rho^2/(8t)}\,G_k(t) \tag{10}$$

with the function $G_k(t)$ given by

$$G_k(t) := \frac{e^{-t/4}}{(4\pi t)^{3/2}} \int_0^\infty \frac{re^{-r^2/(8t)}}{\sinh(r/2)}\cosh(kr)\,\mathrm{d}r.$$

Introducing the function

$$H(t) := \int_0^\infty e^{-\rho^2/(8t)} e^\rho\,\mathrm{d}\rho\,,$$

the bound (10) in combination with (8) yields

$$\sum_{\gamma\in\Gamma} K_k(t;\rho_{\gamma;z}) = O_{\Gamma,K}\big(G_k(t)H(t)\big),$$

where the implied constant equals the implied constant in (8). From this the claim of the proposition follows. □

Corollary 3.4. *For any Fuchsian subgroup* Γ *of the first kind and* $k \in \mathbb{N}_{>0}$*, we have the bound*

$$S_{2k}^{\Gamma}(z) \leq \sum_{\gamma\in\Gamma} K_k(t;\rho_{\gamma;z})$$

for any $t > 0$ *and* $z \in \mathbb{H}$*, where the right-hand side converges uniformly on compact subsets of* M*.*

Proof. Since $k \in \mathbb{N}_{>0}$ and

$$\left|\left(\frac{c\bar{z}+d}{cz+d}\right)^k\left(\frac{z-\gamma\bar{z}}{\gamma z-\bar{z}}\right)^k\right| = 1$$

for any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Gamma$, we deduce for any $t > 0$ and $z \in \mathbb{H}$ from (5) that

$$S_{2k}^{\Gamma}(z) \leq e^{k(k-1)t}\, S_{2k}^{\Gamma}(z) \leq K_k^{\Gamma}(t;z) \leq \sum_{\gamma\in\Gamma} K_k(t;\rho_{\gamma;z}), \tag{11}$$

where the right-hand side of (11) converges uniformly on compact subsets by Proposition 3.3. This proves the claim. □

4 Bounds in the Cocompact Setting

Proposition 4.1. *For any* $\delta > 0$*, there is a constant* $C_\delta > 0$*, such that for any Fuchsian subgroup* Γ *of the first kind and* $k \in \mathbb{N}_{>0}$*, we have the bound*

$$S_{2k}^{\Gamma}(z) \leq k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^{2k}(\rho_{\gamma;z}/2)} + C_\delta\, k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}\geq\delta}} \frac{\rho_{\gamma;z}\, e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)},$$

where we recall that $\rho_{\gamma;z} = \mathrm{dist}_{\mathrm{hyp}}(z,\gamma z)$ *with* $z \in \mathbb{H}$ *and* $\gamma \in \Gamma$*.*

Proof. From Corollary 3.4, we recall for any $t > 0$ and $z \in \mathbb{H}$ the inequality

$$S_{2k}^{\Gamma}(z) \leq \sum_{\gamma \in \Gamma} K_k(t; \rho_{\gamma;z}). \tag{12}$$

We proceed by estimating the right-hand side of (12), i.e., by giving a suitable bound for

$$K_k(t; \rho_{\gamma;z}) = \frac{\sqrt{2}\, e^{-t/4}}{(4\pi t)^{3/2}} \int_{\rho_{\gamma;z}}^{\infty} \frac{r e^{-r^2/(4t)}}{\sqrt{\cosh(r) - \cosh(\rho_{\gamma;z})}}\, T_{2k}\left(\frac{\cosh(r/2)}{\cosh(\rho_{\gamma;z}/2)}\right) \mathrm{d}r\,.$$

We start with some elementary bounds for the Chebyshev polynomials $T_{2k}(X) = \cosh(2k \operatorname{arccosh}(X))$. Using that

$$\operatorname{arccosh}(X) = \log\left(X + \sqrt{X^2 - 1}\right),$$

we find

$$\begin{aligned}
&\operatorname{arccosh}\left(\frac{\cosh(r/2)}{\cosh(\rho_{\gamma;z}/2)}\right) \\
&\quad = \log\left(\frac{1}{\cosh(\rho_{\gamma;z}/2)}\left(\cosh(r/2) + \sqrt{\cosh^2(r/2) - \cosh^2(\rho_{\gamma;z}/2)}\right)\right) \\
&\quad \leq \log\left(\frac{1}{\cosh(\rho_{\gamma;z}/2)}\left(\cosh(r/2) + \sqrt{\cosh^2(r/2) - 1}\right)\right) \\
&\quad = r/2 - \log\left(\cosh(\rho_{\gamma;z}/2)\right).
\end{aligned}$$

Therefore, we obtain the bound

$$T_{2k}\left(\frac{\cosh(r/2)}{\cosh(\rho_{\gamma;z}/2)}\right) = \cosh\left(2k \operatorname{arccosh}\left(\frac{\cosh(r/2)}{\cosh(\rho_{\gamma;z}/2)}\right)\right) \leq \frac{e^{kr}}{\cosh^{2k}(\rho_{\gamma;z}/2)}\,,$$

and hence arrive at

$$\begin{aligned}
S_{2k}^{\Gamma}(z) &\leq \frac{\sqrt{2}\, e^{-t/4}}{(4\pi t)^{3/2}} \sum_{\gamma \in \Gamma} \int_{\rho_{\gamma;z}}^{\infty} \frac{r e^{-r^2/(4t)}}{\sqrt{\cosh(r) - \cosh(\rho_{\gamma;z})}}\, \frac{e^{kr}}{\cosh^{2k}(\rho_{\gamma;z}/2)}\, \mathrm{d}r \\
&= \frac{\sqrt{2}\, e^{-t/4}}{(4\pi t)^{3/2}} \sum_{\gamma \in \Gamma} \frac{1}{\cosh^{2k}(\rho_{\gamma;z}/2)} \int_{\rho_{\gamma;z}}^{\infty} \frac{r e^{-r^2/(4t)+kr}}{\sqrt{\cosh(r) - \cosh(\rho_{\gamma;z})}}\, \mathrm{d}r.
\end{aligned} \tag{13}$$

We next multiply both sides of inequality (13) by $te^{-s(s-1)t}$ with $s \in \mathbb{R}$, $s > k$, and integrate from $t = 0$ to $t = \infty$. Recalling from [GR81], formula 3.325, namely

$$\int_0^\infty e^{-a^2 t} e^{-b^2/(4t)}\, t^{1/2} \frac{\mathrm{d}t}{t} = \frac{\sqrt{\pi}}{a} e^{-ab},$$

we arrive with $a = s - 1/2$ and $b = r$ at the bound

$$\frac{S_{2k}^{\Gamma}(z)}{(s(s-1) - k(k-1))^2} \leq \frac{\sqrt{2\pi}}{(4\pi)^{3/2}(s-1/2)} \sum_{\gamma \in \Gamma} \frac{1}{\cosh^{2k}(\rho_{\gamma;z}/2)} \times \int_{\rho_{\gamma;z}}^{\infty} \frac{re^{-(s-1/2)r+kr}}{\sqrt{\cosh(r) - \cosh(\rho_{\gamma;z})}}\, \mathrm{d}r.$$

Now, let $s = k + 1$, to get

$$S_{2k}^{\Gamma}(z) \leq \frac{\sqrt{2}}{2\pi} \frac{k^2}{k + 1/2} \sum_{\gamma \in \Gamma} \frac{1}{\cosh^{2k}(\rho_{\gamma;z}/2)} \int_{\rho_{\gamma;z}}^{\infty} \frac{re^{-r/2}}{\sqrt{\cosh(r) - \cosh(\rho_{\gamma;z})}}\, \mathrm{d}r. \tag{14}$$

To finish, we will estimate the integral in (14) in a manner similar to the proof of Lemma 4.2 in [JK13]. We start by first considering the case, where $\rho \geq \delta$. Let us then use the decomposition

$$\int_{\rho_{\gamma;z}}^{\infty} \cdots = \int_{\rho_{\gamma;z}}^{\rho_{\gamma;z}+\log(4)} \cdots + \int_{\rho_{\gamma;z}+\log(4)}^{\infty} \cdots$$

For $r \in [\rho_{\gamma;z}, \rho_{\gamma;z} + \log(4)]$, we have the bound

$$\cosh(r) - \cosh(\rho_{\gamma;z}) = (r - \rho_{\gamma;z}) \sinh(r_*) \geq (r - \rho_{\gamma;z}) \sinh(\rho_{\gamma;z}),$$

where $r_* \in [\rho_{\gamma;z}, \rho_{\gamma;z} + \log(4)]$. With this in mind, we have the estimate

$$\int_{\rho_{\gamma;z}}^{\rho_{\gamma;z}+\log(4)} \frac{re^{-r/2}}{\sqrt{\cosh(r) - \cosh(\rho_{\gamma;z})}}\, \mathrm{d}r \leq \frac{(\rho_{\gamma;z} + \log(4))e^{-\rho_{\gamma;z}/2}}{\sqrt{\sinh(\rho_{\gamma;z})}} \int_{\rho_{\gamma;z}}^{\rho_{\gamma;z}+\log(4)} \frac{\mathrm{d}r}{\sqrt{r - \rho_{\gamma;z}}}$$
$$= 2\sqrt{\log(4)}\, \frac{(\rho_{\gamma;z} + \log(4))e^{-\rho_{\gamma;z}/2}}{\sqrt{\sinh(\rho_{\gamma;z})}}. \tag{15}$$

If $r \geq \rho_{\gamma;z} + \log(4)$, we have

$$\frac{\cosh(r)}{2} \geq \frac{\cosh(\rho_{\gamma;z} + \log(4))}{2} \geq \frac{\cosh(\rho_{\gamma;z})\cosh(\log(4))}{2} \geq \cosh(\rho_{\gamma;z}),$$

so then

$$\cosh(r) - \cosh(\rho_{\gamma;z}) \geq \frac{1}{2}\cosh(r) \geq \frac{e^r}{4},$$

hence

$$\int\limits_{\rho_{\gamma;z}+\log(4)}^{\infty} \frac{re^{-r/2}}{\sqrt{\cosh(r) - \cosh(\rho_{\gamma;z})}}\,\mathrm{d}r \leq 2 \int\limits_{\rho_{\gamma;z}+\log(4)}^{\infty} re^{-r}\,\mathrm{d}r = \frac{(\rho_{\gamma;z} + \log(4) + 1)e^{-\rho_{\gamma;z}}}{2}. \tag{16}$$

Combining inequalities (15) and (16), we find for $\rho_{\gamma;z} \geq \delta$ a suitable constant $C_\delta > 0$ depending on δ such that

$$\int\limits_{\rho_{\gamma;z}}^{\infty} \frac{re^{-r/2}}{\sqrt{\cosh(r) - \cosh(\rho_{\gamma;z})}}\,\mathrm{d}r \leq 2\sqrt{\log(4)}\,\frac{(\rho_{\gamma;z} + \log(4))e^{-\rho_{\gamma;z}/2}}{\sqrt{\sinh(\rho_{\gamma;z})}} + \frac{(\rho_{\gamma;z} + \log(4) + 1)e^{-\rho_{\gamma;z}}}{2} \leq C_\delta\,\rho_{\gamma;z}\,e^{-\rho_{\gamma;z}}.$$

From inequality (14), we thus obtain the bound

$$S_{2k}^{\Gamma}(z) \leq k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}<\delta}} \frac{1}{\cosh^{2k}(\rho_{\gamma;z}/2)} \int\limits_{\rho_{\gamma;z}}^{\infty} \frac{re^{-r/2}}{\sqrt{\cosh(r) - \cosh(\rho_{\gamma;z})}}\,\mathrm{d}r + C_\delta\, k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}\geq\delta}} \frac{\rho_{\gamma;z}\,e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)}. \tag{17}$$

In order to estimate the finite sum in (17), we introduce the function

$$h(\rho) := \int\limits_{\rho}^{\infty} \frac{re^{-r/2}}{\sqrt{\cosh(r) - \cosh(\rho)}}\,\mathrm{d}r = -2 \int\limits_{\rho}^{\infty} \sqrt{\cosh(r) - \cosh(\rho)}\,\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{re^{-r/2}}{\sinh(r)}\right)\mathrm{d}r.$$

We have

$$\frac{\mathrm{d}}{\mathrm{d}\rho}h(\rho) = \int_{\rho}^{\infty} \frac{\sinh(\rho)}{\sqrt{\cosh(r)-\cosh(\rho)}} \frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{re^{-r/2}}{\sinh(r)}\right)\mathrm{d}r$$
$$= \int_{\rho}^{\infty} \frac{\sinh(\rho)}{\sqrt{\cosh(r)-\cosh(\rho)}} \frac{re^{-r/2}}{\sinh(r)}\left(\frac{1}{r}-\frac{1}{2}-\coth(r)\right)\mathrm{d}r.$$

Since $\tanh(r) \leq r$, we have that $\coth(r) \geq 1/r$, so then $1/r - 1/2 - \coth(r) \leq -1/2 < 0$, hence the function $h(\rho)$ is monotone decreasing. Therefore, (17) simplifies to

$$S_{2k}^{\Gamma}(z) \leq k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}<\delta}} \frac{1}{\cosh^{2k}(\rho_{\gamma;z}/2)} \int_{0}^{\infty} \frac{re^{-r/2}}{\sqrt{\cosh(r)-1}}\,\mathrm{d}r + C_{\delta}\,k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}\geq\delta}} \frac{\rho_{\gamma;z}\,e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)}\,.$$

Using the fact that $\sinh(r) \geq r$, we have that

$$\int_{0}^{\infty} \frac{re^{-r/2}}{\sqrt{\cosh(r)-1}}\,\mathrm{d}r = \int_{0}^{\infty} \frac{re^{-r/2}}{\sqrt{2}\sinh(r/2)}\,\mathrm{d}r \leq \sqrt{2}\int_{0}^{\infty} e^{-r/2}\,\mathrm{d}r = 2\sqrt{2}.$$

Therefore, we arrive at the bound

$$S_{2k}^{\Gamma}(z) \leq k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^{2k}(\rho_{\gamma;z}/2)} + C_{\delta}\,k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}\geq\delta}} \frac{\rho_{\gamma;z}\,e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)},$$

as claimed. □

Theorem 4.2. *For any Fuchsian subgroup* Γ *of the first kind,* $k \in \mathbb{N}_{>0}$*, and any compact subset* $K \subseteq M$*, we have the bound*

$$\sup_{z\in K}\left(S_{2k}^{\Gamma}(z)\right) = O_{\Gamma,K}(k),$$

where the implied constant depends solely on Γ *and* K*.*

Proof. From Proposition 4.1, we have the bound

$$S_{2k}^{\Gamma}(z) \leq k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^{2k}(\rho_{\gamma;z}/2)} + C_{\delta}\,k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}\geq\delta}} \frac{\rho_{\gamma;z}\,e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)}$$

$$\leq k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^2(\rho_{\gamma;z}/2)} + C_\delta\, k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}\geq\delta}} \frac{\rho_{\gamma;z}\, e^{-\rho_{\gamma;z}}}{\cosh^2(\rho_{\gamma;z}/2)}\,. \tag{18}$$

In order to estimate the first summand in (18), we observe that the sum is finite and hence is a well-defined continuous function on M, which has a maximum $C'_{\Gamma,K,\delta} > 0$ on K, depending solely on Γ, K, and δ. For $z \in K$, we thus have

$$k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^2(\rho_{\gamma;z}/2)} \leq C'_{\Gamma,K,\delta}\, k. \tag{19}$$

To finish, we use the counting function $N(\rho; z)$ defined by (6) and its bound (7). For the second summand in (18), we then find a constant $C''_{\Gamma,K,\delta} > 0$ depending solely on Γ, K, and δ such that

$$\begin{aligned} C_\delta\, k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}\geq\delta}} \frac{\rho_{\gamma;z}\, e^{-\rho_{\gamma;z}}}{\cosh^2(\rho_{\gamma;z}/2)} &\leq 4\, C_\delta\, k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}\geq\delta}} \rho_{\gamma;z}\, e^{-2\rho_{\gamma;z}} \leq C''_{\Gamma,K,\delta}\, k \int_0^\infty \rho\, e^{-2\rho} e^{\rho}\, \mathrm{d}\rho \\ &= C''_{\Gamma,K,\delta}\, k. \end{aligned} \tag{20}$$

Adding up inequalities (19) and (20) yields the claim keeping in mind that δ can be chosen universally. □

Corollary 4.3. *For any cocompact Fuchsian subgroup* Γ *of the first kind and* $k \in \mathbb{N}_{>0}$, *we have the bound*

$$\sup_{z\in M} \left(S^{\Gamma}_{2k}(z)\right) = O_\Gamma(k),$$

where the implied constant depends solely on Γ.

Proof. The proof is an immediate consequence of Theorem 4.2. □

5 Bounds in the Cofinite Setting

Proposition 5.1. *For a cofinite Fuchsian subgroup* Γ *of the first kind and* $k \in \mathbb{N}_{>0}$, *let* $\varepsilon > 0$ *be such that the neighborhoods of area* ε *around the cusps of* M *are disjoint. Assuming that* $0 < \varepsilon < 2\pi/k$, *we have the bound*

$$\sup_{z\in M} \left(S^{\Gamma}_{2k}(z)\right) = O_{\Gamma,\varepsilon}(k),$$

where the implied constant depends solely on Γ *and* ε.

Proof. For a cusp $p \in \mathcal{C}$, we denote by $U_\varepsilon(p)$ the neighborhood of area ε centered at p. By means of the neighborhoods $U_\varepsilon(p)$, we have the compact subset

$$K_\varepsilon := M \setminus \bigcup_{p \in \mathcal{C}} U_\varepsilon(p)$$

of M. We will now estimate the quantity $S_{2k}^{\Gamma}(z)$ for z ranging through K_ε and $U_\varepsilon(p)$ ($p \in \mathcal{C}$), respectively.

In the first case, we obtain from Theorem 4.2 that

$$\sup_{z \in K_\varepsilon} \big(S_{2k}^{\Gamma}(z) \big) = O_{\Gamma, K_\varepsilon}(k),$$

where the implied constant depends solely on Γ and K_ε.

In order to prove the claim in the second case, we may assume without loss of generality that p is the cusp at infinity and the neighborhood $U_\varepsilon(p)$ is given by the strip

$$\mathcal{S}_{1/\varepsilon} := \{ z \in \mathbb{H} \,|\, 0 \leq x < 1,\, y > 1/\varepsilon \}.$$

For a cusp form $f \in \mathcal{S}_{2k}^{\Gamma}$ of weight $2k$ for Γ, we then consider the expression

$$|f(z)|^2\, y^{2k} = \left| \frac{f(z)}{e^{2\pi i z}} \right|^2 \frac{y^{2k}}{e^{4\pi y}}.$$

The function $|f(z)/e^{2\pi i z}|^2$ is subharmonic and bounded in the strip $\mathcal{S}_{1/\varepsilon}$ and, hence, takes its maximum on the boundary

$$\partial \mathcal{S}_{1/\varepsilon} = \{ z \in \mathbb{H} \,|\, 0 \leq x < 1,\, y = 1/\varepsilon \}$$

of $\mathcal{S}_{1/\varepsilon}$, by the strong maximum principle for subharmonic functions. On the other hand, an elementary calculation shows that the function $y^{2k}/e^{4\pi y}$ takes its maximum at

$$y = \frac{k}{2\pi} < \frac{1}{\varepsilon},$$

and is monotone decreasing for $y > k/(2\pi)$. Therefore, we have

$$\sup_{z \in \mathcal{S}_{1/\varepsilon}} \big(|f(z)|^2\, y^{2k} \big) = \sup_{z \in \partial \mathcal{S}_{1/\varepsilon}} \big(|f(z)|^2\, y^{2k} \big).$$

From this we conclude that

$$\sup_{z \in M} \big(S_{2k}^{\Gamma}(z) \big) = \sup_{z \in K_\varepsilon} \big(S_{2k}^{\Gamma}(z) \big) = O_{\Gamma, K_\varepsilon}(k).$$

Since the compact subset K_ε depends only on M, i.e., on Γ, and on ε, the claim of the proposition follows. □

Theorem 5.2. *For a cofinite Fuchsian subgroup Γ of the first kind and $k \in \mathbb{N}_{>0}$, we have the bound*

$$\sup_{z \in M} \big(S_{2k}^{\Gamma}(z)\big) = O_{\Gamma}(k^{3/2}),$$

where the implied constant depends solely on Γ.

Proof. As in the proof of Proposition 5.1, we choose $\varepsilon > 0$ such that the neighborhoods $U_\varepsilon(p)$ of area ε around the cusps $p \in \mathcal{C}$ are disjoint. These neighborhoods give rise to the compact subset

$$K_\varepsilon := M \setminus \bigcup_{p \in \mathcal{C}} U_\varepsilon(p)$$

of M. As before, we will estimate the quantity $S_{2k}^{\Gamma}(z)$ for z ranging through K_ε and $U_\varepsilon(p)$ ($p \in \mathcal{C}$), respectively. As in the proof of Proposition 5.1, we obtain

$$\sup_{z \in K_\varepsilon} \big(S_{2k}^{\Gamma}(z)\big) = O_{\Gamma, K_\varepsilon}(k), \tag{21}$$

where the implied constant depends solely on Γ and K_ε. Since the choice of ε depends only on M, the implied constant depends in the end solely on Γ.

In order to establish the claimed bound for the cuspidal neighborhoods, we distinguish two cases.

(i) If $0 < \varepsilon < 2\pi/k$, the bound for $S_{2k}^{\Gamma}(z)$ in the cuspidal neighborhoods $U_\varepsilon(p)$ ($p \in \mathcal{C}$) is reduced to the bound (21) as in the proof of Proposition 5.1. The proof of the theorem follows in this case.

(ii) If $\varepsilon \geq 2\pi/k$, we have to modify the estimates for $S_{2k}^{\Gamma}(z)$ in the cuspidal neighborhoods $U_\varepsilon(p)$ ($p \in \mathcal{C}$). As before, we may assume without loss of generality that p is the cusp at infinity and the neighborhood $U_\varepsilon(p)$ is given by the strip

$$\mathcal{S}_{1/\varepsilon} := \{z \in \mathbb{H} \,|\, 0 \leq x < 1,\ y > 1/\varepsilon\}.$$

From the argument given in the proof of Proposition 5.1, we find that

$$\sup_{z \in \mathcal{S}_{k/(2\pi)}} \big(S_{2k}^{\Gamma}(z)\big) = \sup_{z \in \partial\mathcal{S}_{k/(2\pi)}} \big(S_{2k}^{\Gamma}(z)\big),$$

where $\mathcal{S}_{k/(2\pi)}$ is the subset of $\mathcal{S}_{1/\varepsilon}$ given by

$$\mathcal{S}_{k/(2\pi)} := \{z \in \mathbb{H} \,|\, 0 \leq x < 1,\ y > k/(2\pi)\}.$$

Therefore, we are reduced to estimate the quantity $S_{2k}^{\Gamma}(z)$ for z ranging through the set

$$\mathcal{S}_{1/\varepsilon} \setminus \mathcal{S}_{k/(2\pi)} = \{z \in \mathbb{H} \,|\, 0 \le x < 1,\ 1/\varepsilon < y \le k/(2\pi)\}.$$

For this, we will use the bound

$$S_{2k}^{\Gamma}(z) \le k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^{2k}(\rho_{\gamma;z}/2)} + C_\delta\, k \sum_{\substack{\gamma\in\Gamma \\ \rho_{\gamma;z}\ge\delta}} \frac{\rho_{\gamma;z}\, e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)} \tag{22}$$

obtained in Proposition 4.1 with an arbitrarily, but fixed chosen $\delta > 0$. By means of the stabilizer subgroup

$$\Gamma_\infty := \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \,\middle|\, n \in \mathbb{Z} \right\}$$

of the cusp at infinity, we can rewrite inequality (22) as

$$\begin{aligned} S_{2k}^{\Gamma}(z) \le k &\sum_{\substack{\gamma\in\Gamma_\infty \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^{2k}(\rho_{\gamma;z}/2)} + C_\delta\, k \sum_{\substack{\gamma\in\Gamma_\infty \\ \rho_{\gamma;z}\ge\delta}} \frac{\rho_{\gamma;z}\, e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)} \\ &+ k \sum_{\substack{\gamma\in\Gamma\setminus\Gamma_\infty \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^{2k}(\rho_{\gamma;z}/2)} + C_\delta\, k \sum_{\substack{\gamma\in\Gamma\setminus\Gamma_\infty \\ \rho_{\gamma;z}\ge\delta}} \frac{\rho_{\gamma;z}\, e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)} . \end{aligned} \tag{23}$$

Using the formula

$$\cosh^2\left(\frac{\mathrm{dist}_{\mathrm{hyp}}(z,w)}{2}\right) = \frac{|z-\bar{w}|^2}{4\,\mathrm{Im}(z)\,\mathrm{Im}(w)} ,$$

the first two summands on the right-hand side of (23) can be bounded as

$$\begin{aligned} k &\sum_{\substack{\gamma\in\Gamma_\infty \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^{2k}(\rho_{\gamma;z}/2)} + C_\delta\, k \sum_{\substack{\gamma\in\Gamma_\infty \\ \rho_{\gamma;z}\ge\delta}} \frac{\rho_{\gamma;z}\, e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)} \\ &\le k(2\sqrt{2} + C_\delta/e) + 2k \sum_{n=1}^{\infty} \frac{2\sqrt{2} + C_\delta/e}{\big((n/2y)^2 + 1\big)^k} . \end{aligned}$$

By an integral test, we have (recalling formula 3.251.2 from [GR81])

$$\sum_{n=1}^{\infty} \frac{1}{\left((n/2y)^2+1\right)^k} \frac{1}{2y} \leq \int_0^{\infty} \frac{1}{\left(1+\eta^2\right)^k}\, \mathrm{d}\eta = \frac{\sqrt{\pi}\,\Gamma(k-1/2)}{2\,\Gamma(k)},$$

which leads to the bound

$$k \sum_{\substack{\gamma\in\Gamma_\infty \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^{2k}(\rho_{\gamma;z}/2)} + C_\delta\, k \sum_{\substack{\gamma\in\Gamma_\infty \\ \rho_{\gamma;z}\geq\delta}} \frac{\rho_{\gamma;z}\, e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)}$$
$$= O\!\left(k\, y \frac{\Gamma(k-1/2)}{\Gamma(k)}\right) = O\!\left(k^{3/2}\right),$$

keeping in mind that $y \leq k/(2\pi)$ and using Stirling's formula.

We now turn to estimate the third summand on the right-hand side of (23). For fixed $z \in \mathcal{S}_{1/\varepsilon} \setminus \mathcal{S}_{k/(2\pi)}$, the sum in question is finite and bounded by the corresponding sum with $k = 1$. Letting z more generally range across the compact subset given by the closure of $\mathcal{S}_{1/\varepsilon}$, the latter sum takes its maximum on that compact set, which depends solely on Γ, ε, and δ. In summary, we obtain

$$k \sum_{\substack{\gamma\in\Gamma\setminus\Gamma_\infty \\ \rho_{\gamma;z}<\delta}} \frac{2\sqrt{2}}{\cosh^{2k}(\rho_{\gamma;z}/2)} = O_\Gamma(k), \tag{24}$$

where the implied constant depends solely on Γ.

We are left to estimate the fourth summand on the right-hand side of (23). Eventually, by shrinking ε, we may assume that we have $\mathrm{Im}(\gamma z) < 1/\varepsilon$ for all $\gamma \in \Gamma \setminus \Gamma_\infty$; this process depends only on Γ. We then find

$$C_\delta\, k \sum_{\substack{\gamma\in\Gamma\setminus\Gamma_\infty \\ \rho_{\gamma;z}\geq\delta}} \frac{\rho_{\gamma;z}\, e^{-\rho_{\gamma;z}}}{\cosh^{2k}(\rho_{\gamma;z}/2)} \leq C_\delta\, k \sum_{\substack{\gamma\in\Gamma\setminus\Gamma_\infty \\ \rho_{\gamma;z}\geq\delta}} \frac{e^{-\rho_{\gamma;z}/2}}{\cosh^{2k}(\rho_{\gamma;z}/2)}$$
$$\leq C_\delta\, k \sum_{\gamma\in\Gamma\setminus\Gamma_\infty} \frac{e^{-\rho_{\gamma;z,\varepsilon}/2}}{\cosh^{2}(\rho_{\gamma;z,\varepsilon}/2)}, \tag{25}$$

where

$$\rho_{\gamma;z,\varepsilon} := \mathrm{dist}_{\mathrm{hyp}}\big(\gamma z, \partial\mathcal{S}_{1/\varepsilon}\big).$$

Using a counting function similar to (6) with a bound similar to (7), the right-hand side of (25) can be bounded as $O_{\Gamma,\varepsilon}(C_\delta\, k)$ with an implied constant depending solely on Γ and ε, hence solely on Γ.

This completes the proof of the theorem. □

6 Bounds for Covers

In this section, we fix a Fuchsian subgroup $\Gamma_0 \subset \mathrm{PSL}_2(\mathbb{R})$ of the first kind with quotient space $M_0 := \Gamma_0\backslash\mathbb{H}$. We then consider subgroups $\Gamma \subseteq \Gamma_0$, which are of finite index. The quotient space $M = \Gamma\backslash\mathbb{H}$ then is a finite degree cover of M_0. Our main goal in this section is to give uniform bounds for the quantity $S_{2k}^{\Gamma}(z)$ depending solely on k and Γ_0.

Theorem 6.1. *Let Γ_0 be a fixed Fuchsian subgroup of $\mathrm{PSL}_2(\mathbb{R})$ of the first kind and $\Gamma \subseteq \Gamma_0$ any subgroup of finite index. For any $k \in \mathbb{N}_{>0}$, we then have the bound*

$$\sup_{z\in M}\big(S_{2k}^{\Gamma}(z)\big) = O_{\Gamma_0}(k^{3/2}),$$

where the implied constant depends solely on Γ_0.

Proof. Denote by $\pi : M \longrightarrow M_0$ the covering map and by $\mathcal{C}_0$ the set of cusps of M_0. As before, we choose $\varepsilon > 0$ such that the neighborhoods $U_\varepsilon(p_0)$ of area ε around the cusps $p_0 \in \mathcal{C}_0$ are disjoint. These neighborhoods give rise to the compact subset

$$K_{0,\varepsilon} := M_0 \setminus \bigcup_{p_0\in\mathcal{C}_0} U_\varepsilon(p_0)$$

of M_0. By means of $K_{0,\varepsilon}$ we obtain the compact subset $K_\varepsilon := \pi^{-1}(K_{0,\varepsilon})$ of M. For z ranging through K_ε, we use Corollary 3.4 to obtain

$$S_{2k}^{\Gamma}(z) \le \sum_{\gamma\in\Gamma} K_k(t;\rho_{\gamma;z}) \le \sum_{\gamma\in\Gamma_0} K_k(t;\rho_{\gamma;z}). \tag{26}$$

The proofs of Proposition 4.1 and Theorem 4.2 with Γ and K_ε replaced by Γ_0 and $K_{0,\varepsilon}$, respectively, now show that the right-hand side of inequality (26) can be uniformly bounded as $O_{\Gamma_0}(k)$, keeping in mind that the choice of ε and, hence of the compact subset $K_{0,\varepsilon}$, depend solely on Γ_0.

We are thus left to bound $S_{2k}^{\Gamma}(z)$ in the neighborhoods of the cusps of M obtained by pulling back the neighborhoods $U_\varepsilon(p_0)$ for $p_0 \in \mathcal{C}_0$ to M. In order to do this, we can again assume that p_0 is the cusp at infinity and $U_\varepsilon(p_0)$ is given as the strip

$$\mathcal{S}_{1,1/\varepsilon} := \big\{z \in \mathbb{H}\,\big|\, 0 \le x < 1,\ y > 1/\varepsilon\big\}.$$

Furthermore, we may also assume that the cusp $p \in \mathcal{C}$ of M lying over the cusp p_0 is also at infinity of ramification index a, say. The pull-back of the neighborhood $U_\varepsilon(p_0)$ to p via π is then modeled by the strip

$$\mathcal{S}_{a,1/\varepsilon} := \big\{z \in \mathbb{H}\,\big|\, 0 \le x < a,\ y > 1/\varepsilon\big\},$$

which contains the strip

$$\mathcal{S}_{a,a/\varepsilon} := \bigl\{z \in \mathbb{H} \bigm| 0 \leq x < a,\ y > a/\varepsilon\bigr\}$$

of area ε. As in the proof of Theorem 5.2, we distinguish two cases.

(i) If $0 < \varepsilon < 2\pi/k$, i.e., $a/\varepsilon > ak/(2\pi)$, we show as in Proposition 5.1 that

$$\sup_{z\in\mathcal{S}_{a,a/\varepsilon}} \bigl(S_{2k}^{\Gamma}(z)\bigr) = \sup_{z\in\partial\mathcal{S}_{a,a/\varepsilon}} \bigl(S_{2k}^{\Gamma}(z)\bigr),$$

and we are reduced to bound $S_{2k}^{\Gamma}(z)$ in the annulus $\mathcal{S}_{a,1/\varepsilon} \setminus \mathcal{S}_{a,a/\varepsilon}$, which will be done below.

(ii) If $\varepsilon \geq 2\pi/k$, i.e., $a/\varepsilon \leq ak/(2\pi)$, we proceed as in the corresponding part of the proof of Theorem 5.2 to find

$$\sup_{z\in\mathcal{S}_{a,ak/(2\pi)}} \bigl(S_{2k}^{\Gamma}(z)\bigr) = \sup_{z\in\partial\mathcal{S}_{a,ak/(2\pi)}} \bigl(S_{2k}^{\Gamma}(z)\bigr),$$

where $\mathcal{S}_{a,ak/(2\pi)}$ is the strip

$$\mathcal{S}_{a,ak/(2\pi)} := \bigl\{z \in \mathbb{H} \bigm| 0 \leq x < a,\ y > ak/(2\pi)\bigr\},$$

which reduces the problem to bound $S_{2k}^{\Gamma}(z)$ to the region $\mathcal{S}_{a,a/\varepsilon} \setminus \mathcal{S}_{a,ak/(2\pi)}$. As in the proof of Theorem 5.2, we next use inequality (23), observing that we now have

$$\Gamma_\infty = \left\{ \begin{pmatrix} 1 & an \\ 0 & 1 \end{pmatrix} \middle|\, n \in \mathbb{Z} \right\}.$$

The first two summands in (23) can be bounded by an obvious adaption as $O(k^{3/2})$ as z ranges through the set $\mathcal{S}_{a,a/\varepsilon} \setminus \mathcal{S}_{a,ak/(2\pi)}$, where we use in particular that $y \leq ak/(2\pi)$. Furthermore, by increasing the range of summation in the sums (24) and (25) by replacing $\Gamma \setminus \Gamma_\infty$ by $\Gamma_0 \setminus \Gamma_{0,\infty}$, the argument given in the proof of Theorem 5.2 shows that the third and fourth summand in (23) can both be bounded as $O_{\Gamma_0}(k)$. All in all, we obtain in case (ii)

$$\sup_{z\in\mathcal{S}_{a,a/\varepsilon}} \bigl(S_{2k}^{\Gamma}(z)\bigr) = O_{\Gamma_0}(k^{3/2}),$$

and we are also in this case reduced to bound $S_{2k}^{\Gamma}(z)$ in the annulus $\mathcal{S}_{a,1/\varepsilon} \setminus \mathcal{S}_{a,a/\varepsilon}$, which we do next.

To this end, we make again use of the estimate (23) with z ranging through $\mathcal{S}_{a,1/\varepsilon} \setminus \mathcal{S}_{a,a/\varepsilon}$. By estimating the third and the fourth summand in (23) as in (24) and (25) with $\Gamma \setminus \Gamma_\infty$ replaced by $\Gamma_0 \setminus \Gamma_{0,\infty}$, respectively, these two summands can be bounded as $O_{\Gamma_0}(k)$. By proceeding as in the proof of Theorem 5.2, the first and the second summand in (23) can be estimated as $O(k^{1/2}/\varepsilon)$ using that $y \leq a/\varepsilon$.

By adding up all the above estimates, the proof of the theorem is complete. □

Remark 6.2. We note that, if in addition to the hypotheses of Theorem 6.1, the fixed Fuchsian subgroup Γ_0 of $\mathrm{PSL}_2(\mathbb{R})$ of the first kind is cocompact and, hence the subgroup $\Gamma \subseteq \Gamma_0$ of finite index is also cocompact, then the proof of Theorem 6.1 in combination with Corollary 4.3 shows that for any $k \in \mathbb{N}_{>0}$, we then have the bound

$$\sup_{z\in M} \big(S_{2k}^{\Gamma}(z)\big) = O_{\Gamma_0}(k),$$

where the implied constant depends solely on Γ_0.

7 Optimality of the Bounds

In this section we show that the bounds obtained in Corollary 4.3 and Theorem 5.2 are optimal, at least in certain cases.

7.1 Optimality in the Cocompact Setting

In order to address optimality in case that the Fuchsian subgroup Γ of the first kind under consideration is cocompact, we assume in addition that Γ does not contain elliptic elements. We then let ω denote the Hodge bundle on M. For k large enough, we then have by the Riemann–Roch theorem that

$$\begin{aligned} d = \dim_{\mathbb{C}}\big(\mathcal{S}_{2k}^{\Gamma}\big) &= \dim_{\mathbb{C}}\big(H^0(M,\omega^{\otimes 2k})\big) = 2k\deg(\omega) + 1 - g \\ &= 2k\,\frac{\mathrm{vol}_{\mathrm{hyp}}(M)}{4\pi} + 1 - g. \end{aligned}$$

From this we derive for k large enough

$$\sup_{z\in M}\big(S_{2k}^{\Gamma}(z)\big)\,\mathrm{vol}_{\mathrm{hyp}}(M) \geq \int\limits_M S_{2k}^{\Gamma}(z)\,\mu_{\mathrm{hyp}}(z) = d = 2k\,\frac{\mathrm{vol}_{\mathrm{hyp}}(M)}{4\pi} + 1 - g.$$

Dividing by $\mathrm{vol}_{\mathrm{hyp}}(M) = 4\pi(g-1)$, yields

$$\sup_{z\in M}\big(S_{2k}^{\Gamma}(z)\big) \geq \frac{2k-1}{4\pi},$$

which shows that the bound obtained in Corollary 4.3 is optimal for k being large enough.

7.2 Optimality in the Cofinite Setting

In this subsection we will show that the bound obtained in Theorem 5.2 in the cofinite setting is optimal in case that $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. For this, let $f \in \mathcal{S}_{2k}^{\Gamma}$ be an L^2-normalized, primitive, Hecke eigenform with Fourier expansion

$$f(z) = \sum_{n=1}^{\infty} \lambda_f(n)\, e^{2\pi i n z}.$$

In [Xia], letting $\varepsilon > 0$, the author proves as the main result the bounds

$$k^{1/2-\varepsilon} \ll_\varepsilon \sup_{z \in M} \big(|f(z)|^2\, y^{2k}\big) \ll_\varepsilon k^{1/2+\varepsilon},$$

with an implied constant depending only on ε. The lower bound, which is of interest for this subsection, is obtained as follows. For fixed $y > 0$, we compute

$$\int_0^1 |f(x+iy)|^2\, y^{2k}\, \mathrm{d}x = \sum_{n=1}^{\infty} |\lambda_f(n)|^2\, y^{2k}\, e^{-4\pi n y} \geq |\lambda_f(1)|^2\, y^{2k}\, e^{-4\pi y}. \tag{27}$$

From [Xia], we then recall the formula

$$|\lambda_f(1)|^2 = \frac{\pi}{2}\, \frac{(4\pi)^{2k}}{\Gamma(2k)}\, \frac{1}{L\big(\mathrm{Sym}^2(f), 1\big)},$$

where $L\big(\mathrm{Sym}^2(f), s\big)$ ($s \in \mathbb{C}$) denotes the symmetric square L-function associated with the primitive Hecke eigenform f, which can be bounded as

$$k^{-\varepsilon} \ll_\varepsilon L\big(\mathrm{Sym}^2(f), 1\big) \ll_\varepsilon k^{\varepsilon}$$

for any $\varepsilon > 0$. Using Stirling's formula, we arrive at the estimate

$$|\lambda_f(1)|^2 \gg_\varepsilon (2k)^{1/2-\varepsilon} \left(\frac{4\pi e}{2k}\right)^{2k}. \tag{28}$$

Using (28), we derive from (27) the lower bound

$$\int_0^1 |f(x+iy)|^2\, y^{2k}\, \mathrm{d}x \gg_\varepsilon (2k)^{1/2-\varepsilon} \left(\frac{2\pi e}{k}\right)^{2k} \frac{y^{2k}}{e^{4\pi y}}. \tag{29}$$

Evaluating (29) at $y = k/(2\pi)$, we thus obtain the claimed lower bound

$$\sup_{z\in M}\left(|f(z)|^2\, y^{2k}\right) \geq \int_0^1 |f(x+iy)|^2\, y^{2k}\,\mathrm{d}x \gg_\varepsilon k^{1/2-\varepsilon}$$

for k large enough with an implied constant depending on the choice of $\varepsilon > 0$.

Let now $\{f_1, \dots, f_d\}$ be an orthonormal basis of $\mathcal{S}_{2k}^\Gamma$ consisting of primitive Hecke eigenforms. Since $d \gg k$, we arrive with $y = k/(2\pi)$ at

$$\sup_{z\in M}\left(S_{2k}^\Gamma(z)\right) \geq \sum_{j=1}^{d} \int_0^1 \left|f_j(x+iy)\right|^2 y^{2k}\,\mathrm{d}x \gg_\varepsilon k^{3/2-\varepsilon}$$

for k large enough with an implied constant depending on the choice of $\varepsilon > 0$.

Acknowledgements Jorgenson acknowledges support from numerous NSF and PSC-CUNY grants. Kramer acknowledges support from the DFG Graduate School *Berlin Mathematical School* and from the DFG International Research Training Group *Moduli and Automorphic Forms*.

References

[AU95] A. Abbes, E. Ullmo, Comparaison des métriques d'Arakelov et de Poincaré sur $X_0(N)$. Duke Math. J. **80**, 295–307 (1995)

[BF13] R. Berman, G. Freixas i Montplet, An arithmetic Hilbert-Samuel theorem for singular Hermitian line bundles and cusp forms. Compos. Math. **150**, 1703–1728 (2014)

[Fay] J. Fay, Fourier coefficients of the resolvent for a Fuchsian group. J. Reine Angew. Math. **293/294**, 143–203 (1977)

[Fischer] J. Fischer, *An Approach to the Selberg Trace Formula via the Selberg Zeta-Function.* Lecture Notes in Mathematics, vol. 1253 (Springer, New York, 1987)

[GR81] I. Gradshteyn, I. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1981)

[HS10] R. Holowinsky, K. Soundararajan, Mass equidistribution for Hecke eigenforms. Ann. Math. **172**, 1517–1528 (2010)

[HT13] G. Harcos, N. Templier, On the sup-norm of Maass cusp forms of large level III. Math. Ann. **356**, 209–216 (2013)

[Javanpeykar] A. Javanpeykar, Polynomial bounds for Arakelov invariants of Belyi curves. With an appendix by P. Bruin. Algebra Number Theory **8**, 89–140 (2014)

[JK04] J. Jorgenson, J. Kramer, Bounding the sup-norm of automorphic forms. Geom. Funct. Anal. **14**, 1267–1277 (2004)

[JK09] J. Jorgenson, J. Kramer, Bounds on Faltings's delta function through covers. Ann. Math. (2) **170**, 1–43 (2009)

[JK13] J. Jorgenson, J. Kramer, Effective bounds for Faltings's delta function. Dedicated to Christophe Soulé at his sixtieth birthday. Ann. Fac. Sci. Toulouse Math. **23**(6), 665–698 (2014)

[JL95] J. Jorgenson, R. Lundelius, Convergence theorems for relative spectral functions on hyperbolic Riemann surfaces of finite volume. Duke Math. J. **80**, 785–819 (1995)

[Kaenel1] R. von Känel, An effective proof of the hyperelliptic Shafarevich conjecture. J. Théor. Nombres Bordeaux **26**, 507–530 (2014)

[Kaenel2] R. von Känel, Integral points on moduli schemes of elliptic curves. Trans. Lond. Math. Soc. **1**, 85–115 (2014)

[Lau] Y.-K. Lau, Equidistribution of Hecke eigenforms on the arithmetic surface $\Gamma_0(N)\backslash\mathbb{H}$. J. Number Theory **96**, 400–416 (2002)

[LS03] W. Luo, P. Sarnak, Mass equidistribution for Hecke eigenforms. Commun. Pure Appl. Math. **56**, 874–891 (2003)

[MU98] P. Michel, E. Ullmo, Points de petite hauteur sur les courbes modulaires $X_0(N)$. Invent. Math. **131**, 645–674 (1998)

[Oshima] K. Oshima, Completeness relations for Maass Laplacians and heat kernels on the super Poincaré upper half-plane. J. Math. Phys. **31**, 3060–3063 (1990)

[Pazuki] F. Pazuki, Remarques sur une conjecture de Lang. J. Théor. Nombres Bordeaux **22**, 161–179 (2010)

[Templier] N. Templier, Large values of modular forms. Cambridge J. Math. **2**, 91–116 (2014)

[Xia] H. Xia, On L^∞-norms of holomorphic cusp forms. J. Number Theory **124**, 400–416 (2007)

Fivebranes and 4-Manifolds

Abhijit Gadde, Sergei Gukov, and Pavel Putrov

Abstract We describe rules for building 2d theories labeled by 4-manifolds. Using the proposed dictionary between building blocks of 4-manifolds and 2d $\mathcal{N} = (0, 2)$ theories, we obtain a number of results, which include new 3d $\mathcal{N} = 2$ theories $T[M_3]$ associated with rational homology spheres and new results for Vafa–Witten partition functions on 4-manifolds. In particular, we point out that the gluing measure for the latter is precisely the superconformal index of 2d (0, 2) vector multiplet and relate the basic building blocks with coset branching functions. We also offer a new look at the fusion of defect lines/walls, and a physical interpretation of the 4d and 3d Kirby calculus as dualities of 2d $\mathcal{N} = (0, 2)$ theories and 3d $\mathcal{N} = 2$ theories, respectively.

CALT 68-2904

1 Introduction

We study a class of 2d $\mathcal{N} = (0, 2)$ theories $T[M_4]$ labeled by 4-manifolds (with boundary) that enjoys all the standard operations on 4-manifolds, such as cutting, gluing, and the Kirby moves [GS99]. Since the world-sheet SCFT of a heterotic string is a prominent member of this class of 2d $\mathcal{N} = (0, 2)$ theories we shall call it "class $\mathcal{H}$" in what follows. By analogy with theories of class $\mathcal{S}$ and class $\mathcal{R}$ that can be thought of as compactifications of six-dimensional (2, 0) theory on 2-manifolds [GMN10, Gai12, AGT10] and 3-manifolds [DGH11, DGG1, CCV], respectively, a theory $T[M_4]$ of class $\mathcal{H}$ can be viewed as the effective two-dimensional theory describing the physics of fivebranes wrapped on a 4-manifold M_4.

If 2d theories $T[M_4]$ are labeled by 4-manifolds, then what are 4-manifolds labeled by? Unlike the classification of 2-manifolds and 3-manifolds that was of great help in taming the zoo of theories $T[M_2]$ and $T[M_3]$, the world of 4-manifolds is much richer and less understood. In particular, the answer to the above question

A. Gadde • S. Gukov (✉) • P. Putrov
California Institute of Technology, Pasadena, CA 91125, USA
e-mail: gukov@theory.caltech.edu

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_7

is not known at present if by a 4-manifold one means a *smooth* 4-manifold. And, not surprisingly, there will be many points in our journey where this richness of the world of 4-manifolds will translate into rich physics of 2d $\mathcal{N} = (0, 2)$ theories $T[M_4]$. We hope that exploring the duality between 4-manifolds and theories $T[M_4]$ sufficiently far will provide insights into classification of smooth structures in dimension four.

In dimensions ≤ 6, every combinatorial manifold—a.k.a. simplicial complex or a manifold with piecewise linear (PL) structure—admits a unique compatible smooth (DIFF) structure. However, not every topological 4-manifold admits a smooth structure:

$$\text{DIFF} \;=\; \text{PL} \;\subset\; \text{TOP} \tag{1}$$

and, furthermore, the smooth structure on a given topological 4-manifold may not be unique (in fact, M_4 can admit infinitely many smooth structures). When developing a dictionary between M_4 and $T[M_4]$, we will use various tools from string theory and quantum field theory which directly or indirectly involve derivatives of various fields on M_4. Therefore, in our duality between M_4 and $T[M_4]$ all 4-manifolds are assumed to be smooth, but not necessarily compact. In particular, it makes sense to ask what the choice of smooth or PL structure on M_4 means for the 2d theory $T[M_4]$, when the 4-manifold admits multiple smooth structures.

Returning to the above question, the basic topological invariants of a (compact) 4-manifold M_4 are the Betti numbers $b_i(M_4)$ or combinations thereof, such as the Euler characteristic and the signature:

$$\begin{aligned} b_2 &= b_2^+ + b_2^- \\ \sigma &= b_2^+ - b_2^- = \frac{1}{3}\int_{M_4} p_1 \\ \chi &= 2 - 2b_1 + b_2^+ + b_2^- \end{aligned} \tag{2}$$

At least in this paper, we will aim to understand fivebranes on simply connected 4-manifolds. In particular, all compact 4-manifolds considered below will have $b_1(M_4) = 0$. We will be forced, however, to deviate from this assumption (in a minimal way) when discussing cutting and gluing, where non-trivial fundamental groups $|\pi_1(M_4)| < \infty$ will show up.

As long as $b_1 = 0$, there are only two non-trivial integer invariants in (2), which sometimes are replaced by the following topological invariants:

$$\begin{aligned} \chi_h(M_4) &= \frac{\chi(M_4) + \sigma(M_4)}{4} \\ c(M_4) &= 2\chi(M_4) + 3\sigma(M_4) \quad (= c_1^2 \text{ when } M_4 \text{ is a complex surface}) \end{aligned} \tag{3}$$

also used in the literature on 4-manifolds. These two integer invariants (or, simply b_2 and σ) determine the rank and the signature of the bilinear intersection form

$$Q_{M_4} : \Gamma \otimes \Gamma \;\to\; \mathbb{Z} \tag{4}$$

on the homology lattice

$$\Gamma \;=\; H_2(M_4;\mathbb{Z})/\mathrm{Tors} \tag{5}$$

The intersection pairing Q_{M_4} (or, simply, Q) is a nondegenerate symmetric bilinear integer-valued form, whose basic characteristics include the rank, the signature, and the parity (or type). While the first two are determined by $b_2(M_4)$ and $\sigma(M_4)$, the type is defined as follows. The form Q is called even if all diagonal entries in its matrix are even; otherwise it is odd. We also define

$$\Gamma^* = H^2(M_4;\mathbb{Z})/\mathrm{Tors} \tag{6}$$

The relation between the two lattices Γ and Γ^* will play an important role in construction of theories $T[M_4]$ and will be discussed in Sect. 2.

For example, the intersection form for the Kümmer surface has a matrix representation

$$E_8 \oplus E_8 \oplus 3\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{7}$$

where $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ is the intersection form for $S^2 \times S^2$ and E_8 is minus the Cartan matrix for the exceptional Lie algebra by the same name. A form Q is called positive (resp. negative) definite if $\sigma(Q) = \mathrm{rank}(Q)$ (resp. $\sigma(Q) = -\mathrm{rank}(Q)$) or, equivalently, if $Q(\gamma,\gamma) > 0$ (resp. $Q(\gamma,\gamma) < 0$) for all non-zero $\gamma \in \Gamma$. There are finitely many unimodular[1] definite forms of a fixed rank. Thus, in the above example the intersection form for $S^2 \times S^2$ is indefinite and odd, whereas E_8 is the unique unimodular negative definite even form of rank 8.

If M_4 is a closed simply connected oriented 4-manifold, its homeomorphism type is completely determined by Q. To be a little more precise, according to the famous theorem of Michael Freedman [Fre82], compact simply connected topological 4-manifolds are completely characterized by an integral unimodular symmetric bilinear form Q and the Kirby–Siebenmann triangulation obstruction invariant $\alpha(M_4) \in H^4(M_4;\mathbb{Z}_2) \cong \mathbb{Z}_2$, such that $\frac{\sigma}{8} \equiv \alpha \mod 2$ if Q is even. In particular, there is a unique topological 4-manifold with the intersection pairing E_8. This manifold, however, does not admit a smooth structure. Indeed, by Rokhlin's theorem, if a simply connected smooth 4-manifold has an even intersection form Q,

[1] That is, $\det Q = \pm 1$.

then $\sigma(M_4)$ is divisible by 16. There is, however, a *non-compact* smooth manifold with E_8 intersection form that will be one of our examples below: it corresponds to a nice 2d theory $T[E_8]$, which for a single fivebrane we propose to be a realization of level-1 E_8 current algebra used in the world-sheet SCFT of a heterotic string [GSW87, Sect. 6] or in the construction of E-strings [MNVW98]:

$$T[E_8] \;=\; \text{(bosonization of) 8 Fermi multiplets} \tag{8}$$

In the case of compact smooth 4-manifolds, the story is a lot more complicated and the complete classification is not known at present. One major result that will be important to us in what follows is the Donaldson's theorem [Don83], which states that the intersection form Q of a smooth simply connected positive (resp. negative) definite 4-manifold is equivalent over integers to the standard diagonal form $\text{diag}(1, 1, \ldots, 1)$ or $\text{diag}(-1, -1, \ldots, -1)$, respectively. (This result applies to compact M_4 and admits a generalization to 4-manifolds bounded by homology spheres, which we will also need in the study of 2d theories $T[M_4]$.) In particular, since $E_8 \oplus E_8$ is not diagonalizable over integers, the unique topological 4-manifold with this intersection form does not admit a smooth structure.[2] Curiously, this, in turn, implies that $\mathbb{R}^4$ does not have a unique differentiable structure.

We conclude this brief introduction to the wild world of 4-manifolds by noting that any non-compact topological 4-manifold admits a smooth structure [Qui82]. In fact, an interesting feature of non-compact 4-manifolds considered in this paper—that can be viewed either as a good news or as a bad news—is that they all admit *uncountably* many smooth structures.

In order to preserve supersymmetry in two remaining dimensions, the 6d theory must be partially "twisted" along the M_4. The standard way to achieve this is to combine the Euclidean $Spin(4)$ symmetry of the 4-manifold with (part of) the R-symmetry. Then, different choices—labeled by homomorphisms from $Spin(4)$ to the R-symmetry group, briefly summarized in Appendix 1—lead to qualitatively different theories $T[M_4]$, with different amount of supersymmetry in two dimensions, etc. The choice we are going to consider in this paper is essentially (the 6d lift of) the topological twist introduced by Vafa and Witten [VW94], which leads to $(0, 2)$ supersymmetry in two dimensions. In fact, the partition function of the Vafa–Witten TQFT that, under certain conditions, computes Euler characteristics of instanton moduli spaces also plays an important role in the dictionary between 4-manifolds and the corresponding 2d $\mathcal{N} = (0, 2)$ theories $T[M_4]$.

The basic "protected quantity" of any two-dimensional theory with at least $\mathcal{N} = (0, 1)$ supersymmetry is the elliptic genus [Wit87] defined as a partition function on a 2-torus T^2 with periodic (Ramond) boundary conditions for fermions. In the present case, it carries information about all left-moving states of the 2d $\mathcal{N} = (0, 2)$ theory $T[M_4]$ coupled to the supersymmetric Ramond ground states from the right.

[2]Note, this cannot be deduced from the Rokhlin's theorem as in the case of the E_8 manifold.

To be more precise, we shall consider the "flavored" version of the elliptic genus (studied in this context, e.g., in [GGP13, BEHT13]),

$$\mathcal{I}_{T[M_4]}(q,x) \ := \ \mathrm{Tr}_{\mathcal{H}}(-1)^F q^{L_0} x^f \,, \tag{9}$$

that follows the standard definition of the superconformal index in radial quantization and carries extra information about the flavor symmetry charges f. In general, the flavor symmetry group of $T[M_4]$ is $U(1)^{b_2} \times G_{3d}$, where the second factor is associated with the boundary $M_3 = \partial M_4$ and is gauged upon gluing operations. Defined as a supersymmetric partition function on a torus T^2 with a modular parameter τ (where, as usual, $q = e^{2\pi i \tau}$), the index $\mathcal{I}_{T[M_4]}(q;x)$ has a nice interpretation as an invariant of the 4-manifold computed by the topological theory on M_4.

Indeed, since the theory $T[M_4]$ was obtained by compactification from six dimensions on a 4-manifold, its supersymmetric partition function on a torus can be identified with the partition function of the 6d $(2,0)$ theory on $T^2 \times M_4$. As usual, by exchanging the order of compactification, we obtain two perspectives on this fivebrane partition function

6d $(2,0)$ theory
on $T^2 \times M_4$

$\swarrow$ $\searrow$

$\mathcal{N} = 4$ super-Yang-Mills on M_4 | 2d $(0,2)$ theory $T[M_4]$ on T^2

that are expected to produce the same result. If we compactify first on M_4, we obtain a 2d theory $T[M_4]$, whose partition function on T^2 is precisely the flavored elliptic genus (9). On the other hand, if we first compactify on T^2, we get $\mathcal{N} = 4$ super-Yang-Mills[3] with the Vafa–Witten twist on M_4 and coupling constant τ. This suggests the following natural relation:

$$Z^G_{\mathrm{VW}}[M_4](q,x) \ = \ \mathcal{I}_{T[M_4;G]}(q,x) \tag{10}$$

that will be one of our main tools in matching 4-manifolds with 2d $\mathcal{N} = (0,2)$ theories $T[M_4]$. Note, this in particular requires M_4 to be a smooth 4-manifold. Both sides of (10) are known to exhibit nice modular properties under certain favorable assumptions [VW94, Wit87] that we illustrate in numerous examples below.

In this paper, we approach the correspondence between 4-manifolds and 2d $\mathcal{N} = (0,2)$ theories $T[M_4]$ mainly from the viewpoint of cutting and gluing.

[3] Sometimes, to avoid clutter, we suppress the choice of the gauge group, G, which in most of our applications will be either $G = U(N)$ or $G = SU(N)$ for some $N \geq 1$. It would be interesting to see if generalization to G of Cartan type D or E leads to new phenomena. We will not aim to do this analysis here.

For this reason, not only 4-manifolds with boundary are unavoidable, but they are also the main subject of interest. As a result, interesting new phenomena, such as a generalization of the Freed–Witten anomaly [FW99] to manifolds with boundary, come into play. It also affects the relation (10), where the left-hand side naturally becomes a function of boundary conditions, and leads to one interesting novelty discussed in Sect. 3.10. Namely, in order to interpret the Vafa–Witten partition function on a non-compact 4-manifold as the index (9), it is convenient to make a certain transformation—somewhat akin to a change of basis familiar in the literature on the superconformal index [GRRY11]—changing *discrete* labels associated with boundary conditions to *continuous* variables.

The type of the topological twist that leads to 2d $(0,2)$ theory $T[M_4]$, namely the Vafa–Witten twist, can be realized on the world-volume of fivebranes wrapped on a coassociative submanifold M_4 inside a seven-dimensional manifold with G_2 holonomy [BVS95, BT96]. Locally, in the vicinity of M_4, this seven-dimensional manifold always looks like the bundle of self-dual 2-forms over M_4 (see, e.g., [AG04] for a pedagogical review). This realization of the 6d $(2,0)$ theory on the world-volume of M-theory fivebranes embedded in 11d space-time can provide some useful clues about the 2d superconformal theory $T[M_4]$, especially when the number of fivebranes is large, $N \gg 1$, and the system admits a holographic dual supergravity description (cf. Appendix 1 for a brief survey).

In the case of fivebranes on coassociative 4-manifolds, the existence of the holographic dual supergravity solution [GKW00, GK02, BB13] requires M_4 to admit a conformally half-flat structure, i.e., metric with anti-self-dual Weyl tensor. Since the signature of the 4-manifold can be expressed as the integral

$$\sigma(M_4) \;=\; \frac{1}{12\pi^2}\int_{M_4}\left(|W_+|^2 - |W_-|^2\right) \tag{11}$$

where $W_\pm$ are the self-dual and anti-self-dual components of the Weyl tensor, it suggests to focus on 2d $\mathcal{N} = (0,2)$ superconformal theories $T[M_4]$ associated with negative definite M_4. As we explained earlier, negative definite 4-manifolds are very simple in the smooth category and, curiously, $W_+ = 0$ also happens to be the condition under which instantons on M_4 admit a description [AW77] that involves holomorphic vector bundles (on the twistor space of M_4), monads, and other standard tools from $(0,2)$ model building.

The holographic dual and the anomaly of the fivebrane system also allow to express left and right moving central charges of the 2d $\mathcal{N} = (0,2)$ superconformal theory $T[M_4]$ via basic topological invariants (2) of the 4-manifold. Thus, in the case of the 6d $(2,0)$ theory of type G one finds [BB13, ABT10]:

$$\begin{aligned} c_R &= \frac{3}{2}(\chi+\sigma)r_G + (2\chi+3\sigma)d_G h_G \\ c_L &= \chi r_G + (2\chi+3\sigma)d_G h_G \end{aligned} \tag{12}$$

Table 1 The dictionary between geometry and physics

4-Manifold M_4	2d (0, 2) theory $T[M_4]$
Handle slides	Dualities of $T[M_4]$
Boundary conditions	Vacua of $T[M_3]$
3d Kirby calculus	Dualities of $T[M_3]$
Cobordism	Domain wall (interface)
From M_3^- to M_3^+	Between $T[M_3^-]$ and $T[M_3^+]$
Gluing	Fusion
Vafa–Witten	Flavored (equivariant)
Partition function	Elliptic genus
Z_{VW}(cobordism)	Branching function
Instanton number	L_0
Embedded surfaces	Chiral operators
Donaldson polynomials	Chiral ring relations

where $r_G = \text{rank}(G)$, $d_G = \dim(G)$, and h_G is the Coxeter number. In particular, for a single fivebrane ($r_G = 1$ and $d_G h_G = 0$) these expressions give $c_L = \chi$ and $c_R = 3 + 3b_2^+$, suggesting that b_2^- is the number of Fermi multiplets[4] in the 2d $\mathcal{N} = (0, 2)$ theory $T[M_4; U(1)]$. This conclusion agrees with the direct counting of bosonic and fermionic Kaluza–Klein modes [Gan96] and confirms (8). As we shall see in the rest of this paper, the basic building blocks of 2d theories $T[M_4]$ are indeed very simple and, in many cases, can be reduced to Fermi multiplets charged under global flavor symmetries (that are gauged in gluing operations). However, the most interesting part of the story is about operations on 2d (0, 2) theories that correspond to *gluing*.

The paper is organized as follows. In Sect. 2 we describe the general ideas relating 4-manifolds and the corresponding theories $T[M_4]$, fleshing out the basic elements of the dictionary in Table 1. Then, we study the proposed rules in more detail and present various tests as well as new predictions for Vafa–Witten partition functions on 4-manifolds (in Sect. 3) and for 2d walls and boundaries in 3d $\mathcal{N} = 2$ theories (in Sect. 4).

The relation between Donaldson invariants of M_4 and $\overline{\mathcal{Q}}_+$-cohomology of the corresponding 2d (0, 2) theory $T[M_4]$ will be discussed elsewhere. More generally, and as we already remarked earlier, it would be interesting to study to what extent $T[M_4]$, viewed as an invariant of 4-manifolds, can detect smooth structures. In particular, it would be interesting to explore the relation between $T[M_4]$ and other invariants of smooth 4-manifolds originating from physics, such as the celebrated Seiberg–Witten invariants [SW94, Wit94] or various attempts based on gravity [Roh89, Ass96, Pfe04, Sla09].

[4] Recall, that a free Fermi multiplet contributes to the central charge $(c_L, c_R) = (1, 0)$.

2 2d Theories Labeled by 4-Manifolds

Building theories $T[M_4]$ in many ways follows the same set of rules and tricks as building 4-manifolds. Here, we describe some of the basic operations in the world of 4-manifolds and propose their realization in the world of supersymmetric gauge theories. While the emphasis is certainly on explaining the general rules, we supplement each part with concrete examples and/or new calculations. More examples, with further details, and new predictions based on the proposed relations in Table 1 will be discussed in Sects. 3 and 4.

2.1 Kirby Diagrams and Plumbing

We start by reviewing the standard construction of 4-manifolds, based on a handle decomposition, mostly following [GS99] (see also [Akb12]). Thus, if M_4 is connected, we take a single 0-handle ($\cong D^4$) and successively attach to it k-handles ($\cong D^k \times D^{4-k}$) with $k = 1, 2, 3$. Then, depending on the application in mind, we can either stop at this stage (if we are interesting in constructing non-compact 4-manifolds) or cap it off with a 4-handle ($\cong D^4$) if the goal is to build a compact 4-manifold.

The data associated with this process is usually depicted in the form of a *Kirby diagram*, on which every k-handle ($\cong D^k \times D^{4-k}$) is represented by its attaching region, $S^{k-1} \times D^{4-k}$, or by its attaching sphere, S^{k-1}. To be a little more precise, a Kirby diagram of a smooth connected 4-manifold M_4 usually shows only 1-handles and 2-handles because 3-handles and 4-handles attach essentially in a unique way [LP72]. Moreover, in our applications we typically will not see 1-handles either (due to our intention to work with simply connected 4-manifolds). Indeed, regarding a handle decomposition of M_4 as a cell complex, its k-th homology group becomes an easy computation in which k-handles give rise to generators and $(k+1)$-handles give rise to relations. The same interpretation of the handlebody as a cell complex can be also used for the computation of the fundamental group, where 1-handles correspond to generators and 2-handles lead to relations. Therefore, the easiest way to ensure that M_4 is simply connected is to avoid using 1-handles at all.

Then, for this class of 4-manifolds, Kirby diagrams only contain framed circles, i.e., attaching spheres of 2-handles, that can be knotted and linked inside S^3 (= boundary of the 0-handle). To summarize, we shall mostly work with 4-manifolds labeled by framed links in a 3-sphere,

$$M_4 \quad : \qquad K_1^{a_1}\, K_2^{a_2}\, \ldots\, K_n^{a_n} \tag{13}$$

where K_i denotes the i-th component of the link and $a_i \in \mathbb{Z}$ is the corresponding framing coefficient. Examples of Kirby diagrams for simple 4-manifolds are shown in Figs. 1, 2, and 3.

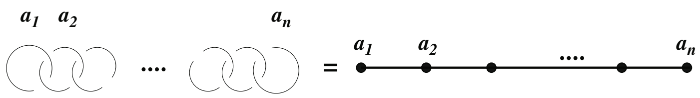


Fig. 1 A Kirby diagram and the corresponding plumbing graph for the plumbing 4-manifold associated with the string $(a_1, a_2, \ldots, a_n)$

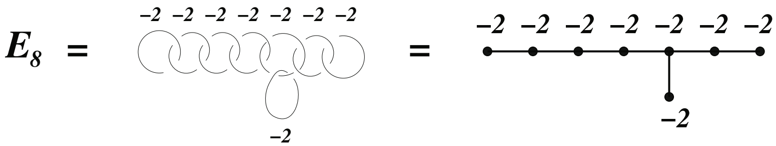
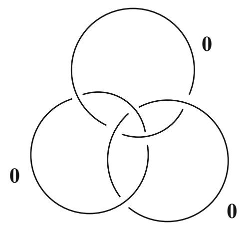


Fig. 2 A Kirby diagram and the corresponding plumbing graph for the E_8 manifold with $b_2 = -\sigma = 8$ and $\partial E_8 \approx \Sigma(2, 3, 5)$

Fig. 3 Kirby diagram of a 4-manifold bounded by a 3-torus T^3



At this stage, it is important to emphasize that Kirby diagrams are not quite unique: there are certain moves which relate different presentations of the same 4-manifold. We refer the reader to excellent monographs [GS99, Akb12] on Kirby calculus, of which most relevant to us is the basic tool called 2-handle slide. Indeed, since our assumptions led us to consider 4-manifolds built out of 2-handles,[5] occasionally we will encounter the operation of sliding a 2-handle i over a 2-handle j. It changes the Kirby diagram and, in particular, the framing coefficients:

$$\begin{aligned} a_j &\mapsto a_i + a_j \pm 2\text{lk}(K_i, K_j) \\ a_i &\mapsto a_i \end{aligned} \tag{14}$$

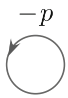
where the sign depends on the choice of orientation ("+" for handle addition and "−" for handle subtraction) and $\text{lk}(K_i, K_j)$ denotes the linking number. We will see in what follows that this operation corresponds to changing the basis of flavor charges.

In the class of non-compact simply connected 4-manifolds (13) labeled by framed links, the simplest examples clearly correspond to Kirby diagrams where all

[5]Another nice property of such 4-manifolds is that they admit an *achiral* Lefschetz fibration over the disk [Har79].

K_i are copies of the unknot. Many[6] such 4-manifolds can be equivalently represented by graphs with integer "weights" assigned to the vertices, somewhat similar to quiver diagrams that conveniently encode the spectrum of fields and interactions in a large class of gauge theories. The 4-manifolds in question are constructed by gluing together n copies of disk bundles over 2-spheres, $D_i^2 \to S_i^2$, each labeled by an integer Euler class $a_i \in \mathbb{Z}$. Switching the role of the base and the fiber in the gluing process, one builds a simply connected 4-manifold M_4, called *plumbing*, whose handle decomposition involves n two-handles (besides the "universal" 0-handle at the bottom). As usual, we represent such 4-manifolds by Kirby diagrams drawing the attaching framed circles K_i of 2-handles inside S^3.

The simplest non-trivial plumbing manifold corresponds to the Kirby diagram:



(15)

In other words, its handlebody decomposition contains only one 2-handle with framing $-p$, and the resulting manifold M_4 is a twisted D^2 bundle over S^2 or, as a complex manifold, the total space of the $\mathcal{O}(-p)$ bundle over $\mathbb{C}\mathbf{P}^1$,

$$M_4 : \quad \mathcal{O}(-p) \to \mathbb{C}\mathbf{P}^1 \tag{16}$$

For $p > 0$, which we are going to assume in what follows, M_4 is a negative definite plumbing manifold bounded by the Lens space $L(p, 1)$.

Another, equivalent way to encode the same data is by a plumbing graph Υ. In this presentation, each attaching circle K_i of a 2-handle is replaced by a vertex with an integer label a_i, and an edge between two vertices i and j indicates that the corresponding attaching circles K_i and K_j are linked. Implicit in the plumbing graph is the orientation of edges, which, unless noted otherwise, is assumed to be such that all linking numbers are $+1$. More generally, one can consider plumbings of twisted D^2 bundles over higher-genus Riemann surfaces, see, e.g., [Akb12, Sect. 2.1], in which case vertices of the corresponding plumbing graphs are labeled by Riemann surfaces (not necessarily orientable) in addition to the integer labels a_i. However, such 4-manifolds typically have non-trivial fundamental group and we will not consider these generalizations here, focusing mainly on plumbings of 2-spheres.

The topology of a 4-manifold M_4 constructed via plumbing of 2-spheres is easy to read off from its Kirby diagram or the corresponding plumbing graph. Specifically, M_4 is a non-compact simply connected 4-manifold, and one can think of K_i as generators of $\Gamma = H_2(M_4; \mathbb{Z})$ with the intersection pairing

[6] But not all! See Fig. 3 for an instructive (counter)example.

$$Q_{ij} = \begin{cases} \mathrm{lk}(K_i, K_j), & \text{if } i \neq j \\ a_i, & \text{if } i = j \end{cases} \tag{17}$$

For example, the Kirby diagram in Fig. 1 corresponds to

$$Q = \begin{pmatrix} a_1 & 1 & 0 & \cdots & 0 \\ 1 & a_2 & 1 & & \vdots \\ 0 & 1 & & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 & a_n \end{pmatrix} \tag{18}$$

A further specialization to $(a_1, a_2, \ldots, a_n) = (-2, -2, \ldots, -2)$ for obvious reasons is usually referred to as A_n, whereas that in Fig. 2 is called E_8.

Similarly, given a weighted graph Υ, one can plumb disk bundles with Euler numbers a_i over 2-spheres together to produce a 4-manifold $M_4(\Upsilon)$ with boundary $M_3(\Upsilon) = \partial M_4(\Upsilon)$, such that

$$b_1(M_4) = b_1(\Upsilon) \tag{19a}$$

$$b_2(M_4) = \#\{\text{vertices of } \Upsilon\} \tag{19b}$$

In particular, aiming to construct simply connected 4-manifolds, we will avoid plumbing graphs that have loops or self-plumbing constructions. Therefore, in what follows we typically assume that Υ is a tree, relegating generalizations to future work. Besides the basic topological invariants (19), the plumbing tree Υ also gives a nice visual presentation of the intersection matrix $Q(\Upsilon) = (Q_{ij})$, which in the natural basis of $H_2(M_4; \mathbb{Z})$ has entries

$$Q_{ij} = \begin{cases} a_i, & \text{if } i = j \\ 1, & \text{if } i \text{ is connected to } j \text{ by an edge} \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

The eigenvalues and the determinant of the intersection form Q can be also easily extracted from Υ by using the algorithm described below in (32) and illustrated in Fig. 4.

Note, this construction of non-compact 4-manifolds admits vast generalizations that do not spoil any of our assumptions (including the simple connectivity of M_4). Thus, in a Kirby diagram of an arbitrary plumbing tree, we can replace every framed unknot (= attaching circle of a 2-handle) by a framed knot, with a framing coefficient a_i. This does not change the homotopy type of the 4-manifold, but does

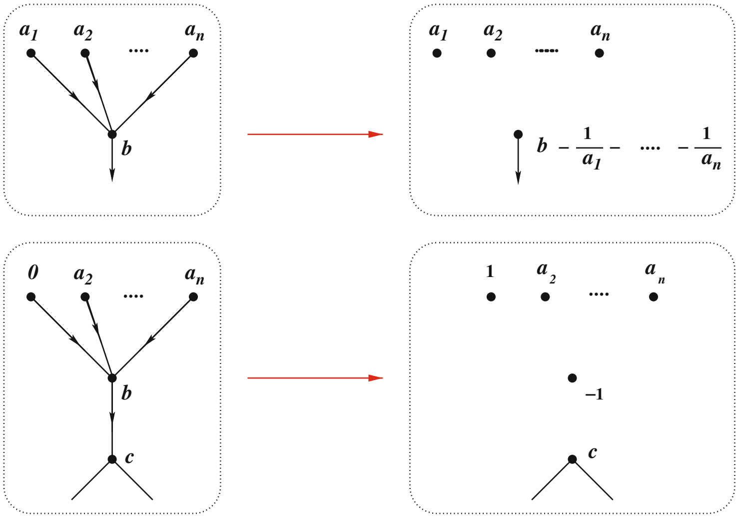


Fig. 4 For a plumbing tree, the eigenvalues (and, therefore, the determinant) of the intersection form Q can be computed by orienting the edges toward a single vertex and then successively eliminating them using the two rules shown here

affect the boundary $M_3 = \partial M_4$. Put differently, all the interesting information about the knot can only be seen at the boundary.

Another important remark is that, although the description of 4-manifolds via plumbing graphs is very nice and simple, it has certain limitations that were already mentioned in the footnote 6. Indeed, if the 4-manifold has self-plumbings or Υ has loops, it may not be possible to consistently convert the Kirby diagram into a plumbing graph without introducing additional labels. An example of such Kirby diagram is shown in Fig. 3, where each pair of the attaching circles K_i with framing $a_i = 0$ has linking number zero. The corresponding 4-manifold, however, is different from that associated with three unlinked copies of the unknot (with plumbing graph that has three vertices and no edges) and the same values of framing coefficients.

Finally, we point out that, since all 4-manifolds constructed in this section have a boundary $M_3 = \partial M_4$, the corresponding 2d $\mathcal{N} = (0,2)$ theory $T[M_4]$ that will be described below should properly be viewed as a boundary condition for the 3d $\mathcal{N} = 2$ theory $T[M_3]$. For example, the plumbing on A_n has the Lens space boundary $M_3 = L(n+1, n)$, while the plumbing on E_8 has the Poincaré sphere boundary $M_3 = \Sigma(2,3,5)$, where

$$\Sigma(a,b,c) := S^5 \cap \{(x,y,z) \in \mathbb{C}^3 \mid x^a + y^b + z^c = 0\} \qquad (21)$$

is the standard notation for a family of Brieskorn spheres. This remark naturally leads us to the study of boundaries M_3 and the corresponding theories $T[M_3]$ for more general sphere plumbings and 4-manifolds (13) labeled by framed links.

2.2 $T[M_4]$ as a Boundary Condition

Since we want to build 4-manifolds as well as the corresponding theories $T[M_4]$ by gluing basic pieces, it is important to develop the physics-geometry dictionary for manifolds with boundary, which will play a key role in gluing and other operations.

2.3 Vacua of the 3d $\mathcal{N} = 2$ Theory $T[M_3]$

Our first goal is to describe supersymmetric vacua of the 3d $\mathcal{N} = 2$ theory $T[M_3]$ associated with the boundary[7] of the 4-manifold M_4,

$$M_3 = \partial M_4 \tag{22}$$

This relation between M_3 and M_4 translates into the statement that 2d $\mathcal{N} = (0, 2)$ theory $T[M_4]$ is a boundary theory for the 3d $\mathcal{N} = 2$ theory $T[M_3]$ on a half-space $\mathbb{R}_+ \times \mathbb{R}^2$. In order to see this, it is convenient to recall that both theories $T[M_3]$ and $T[M_4]$ can be defined as fivebrane configurations (or, compactifications of 6d $(2, 0)$ theory) on the corresponding manifolds, M_3 and M_4. This gives a *coupled* system of 2d-3d theories $T[M_4]$ and $T[M_3]$ since both originate from the same configuration in six dimensions, which looks like $M_3 \times \mathbb{R}_+ \times \mathbb{R}^2$ near the boundary and $M_4 \times \mathbb{R}^2$ away from the boundary. In other words, a 4-manifold M_4 with a boundary M_3 defines a half-BPS (B-type) boundary condition in a 3d $\mathcal{N} = 2$ theory $T[M_3]$.

Therefore, in order to understand a 2d theory $T[M_4]$ we need to identify a 3d theory $T[M_3]$ or, at least, its necessary elements.[8] One important characteristic of a 3d $\mathcal{N} = 2$ theory $T[M_3]$ is the space of its supersymmetric vacua, either in flat

---

[7]Depending on the context, sometimes M_3 will refer to a single component of the boundary.

[8]While this problem has been successfully solved for a large class of 3-manifolds [DGG1, CCV, DGG2], unfortunately it will not be enough for our purposes here and we need to resort to matching M_3 with $T[M_3]$ based on identification of vacua, as was originally proposed in [DGH11]. One reason is that the methods of *loc. cit.* work best for 3-manifolds with sufficiently large boundary and/or fundamental group, whereas in our present context M_3 is itself a boundary and, in many cases, is a rational homology sphere. As we shall see below, 3d $\mathcal{N} = 2$ theories $T[M_3]$ seem to be qualitatively different in these two cases; typically, they are (deformations of) superconformal theories in the former case and massive 3d $\mathcal{N} = 2$ theories in the latter. Another, more serious issue is that 3d theories $T[M_3]$ constructed in [DGG1] do not account for *all* flat connections on M_3, which will be crucial in our applications below. This second issue can be avoided by considering larger 3d theories $T^{(\mathrm{ref})}[M_3]$ that have to do with refinement/categorification and mix all branches

space-time $\mathbb{R}^3$, or on a circle, i.e., in space-time $S^1 \times \mathbb{R}^2$. This will be the subject of our discussion here.

Specifically, when 3d $\mathcal{N} = 2$ theory $T[M_3; G]$ is considered on a circle, its supersymmetric ground states are in one-to-one correspondence with gauge equivalence classes of flat $G_{\mathbb{C}}$ connections on M_3 [DGH11]:

$$d\mathcal{A} + \mathcal{A} \wedge \mathcal{A} \;=\; 0 \tag{23}$$

This follows from the duality between fivebranes on S^1 and D4-branes combined with the fact that D4-brane theory is partially twisted along the 3-manifold M_3. The partial twist in the directions of M_3 is the dimensional reduction of the Vafa–Witten twist [VW94] as well as the GL twist [KW07] of the $\mathcal{N} = 4$ super-Yang-Mills in four dimensions. The resulting $\mathcal{N}_T = 4$ three-dimensional topological gauge theory on M_3 is the equivariant version of the Blau–Thompson theory [BT96, BT97] that localizes on solutions of (23), where $\mathcal{A} = A + iB$ is the Lie$(G_{\mathbb{C}})$-valued connection.

From the viewpoint of the topological Vafa–Witten theory on M_4, solutions to Eq. (23) provide boundary conditions for PDEs in four dimensions. To summarize,

boundary conditions on M_4	$\longleftrightarrow$	complex flat connections on M_3	$\longleftrightarrow$	vacua of $T[M_3]$

In general, complex flat connections on M_3 are labeled by representations of the fundamental group $\pi_1(M_3)$ into $G_{\mathbb{C}}$, modulo conjugation,

$$\mathcal{V}_{T[M_3;G]} \;=\; \text{Rep}\left(\pi_1(M_3) \to G_{\mathbb{C}}\right)/\text{conj.} \tag{24}$$

In particular, in the basic case of abelian theory (i.e., a single fivebrane), the vacua of the 3d $\mathcal{N} = 2$ theory $T[M_3]$ are simply abelian representations of $\pi_1(M_3)$, i.e., elements of $H_1(M_3)$. In the non-abelian case, $G_{\mathbb{C}}$ flat connection on M_3 is described by nice algebraic equations, which play an important role in complex Chern–Simons theory and its relation to quantum group invariants [Guk05].

As will become clear shortly, for many simply connected 4-manifolds (13) built from 2-handles—such as sphere plumbings represented by trees (i.e., graphs without loops)—the boundary M_3 is a rational homology sphere ($b_1(M_3) = 0$) in which case the theory $T[M_3; U(1)]$ has finitely many isolated vacua,

$$\#\{\text{vacua of } T[M_3; U(1)]\} \;=\; |H_1(M_3; \mathbb{Z})| \tag{25}$$

Therefore, the basic piece of data that characterizes $M_3 = \partial M_4$ and the corresponding 3d theory $T[M_3]$ is the first homology group $H_1(M_3; \mathbb{Z})$. Equivalently, when $H_1(M_3; \mathbb{Z})$ is torsion, by the Universal Coefficient Theorem we can label

of flat connections [FGSA, FGP13]. Pursuing this approach should lead to new relations with rich algebraic structure and functoriality of knot homologies.

the vacua of $T[M_3; U(1)]$ by elements of $H^2(M_3; \mathbb{Z})$. Indeed, given a 1-cycle μ in M_3, the Poincaré dual class $[\mu] \in H^2(M_3; \mathbb{Z})$ can be interpreted as the first Chern class $c_1(\mathcal{L}) = [\mu]$ of a complex line bundle $\mathcal{L}$, which admits a flat connection whenever the first Chern class is torsion. The (co)homology groups of the boundary 3-manifold M_3—that, according to (25), determine the vacua of $T[M_3]$—are usually easy to read off from the Kirby diagram of M_4.

Now, once we explained the physical role of the boundary $M_3 = \partial M_4$, we need to discuss its topology in more detail that will allow us to describe complex flat connections on M_3 and, therefore, determine the vacua of the 3d $\mathcal{N} = 2$ theory $T[M_3]$. In general, the boundary of a simply connected 4-manifold (13) labeled by a framed link is an integral surgery on that link in S^3. This operation consists of removing the tubular neighborhood $N(K_i) \cong S^1 \times D^2$ of each link component and then gluing it back in a different way, labeled by a non-trivial self-diffeomorphism $\phi : T^2 \to T^2$ of the boundary torus $\partial N(K_i) \cong T^2$.

This description of the boundary 3-manifold M_3 is also very convenient for describing complex flat connections. Namely, from the viewpoint of T^2 that divides M_3 into two parts, complex flat connections on M_3 are those which can be simultaneously extended from the boundary torus to $M_3 \setminus K_i$ and $N(K_i) \cong S^1 \times D^2$, equivalently, the intersection points

$$\mathcal{V}_{T[M_3]} \; = \; \mathcal{V}_{T[M_3 \setminus K]} \cap \phi \left(\mathcal{V}_{T[S^1 \times D^2]} \right) \tag{26}$$

Here, the representation varieties of the knot complement and the solid torus can be interpreted as (A, B, A) branes in the moduli space of G Higgs bundles on T^2. In this interpretation, ϕ acts as an autoequivalence on the category of branes, see, e.g., [Guk07] for some explicit examples and the computation of (26) in the case $G_{\mathbb{C}} = SL(2, \mathbb{C})$.

Coming back to the vacua (25), the cohomology group $H^2(M_3; \mathbb{Z})$ can be easily deduced from the long exact sequence for the pair (M_4, M_3) with integer coefficients:

$$\begin{array}{ccccccccc} 0 \to H^2(M_4, M_3) & \to & H^2(M_4) & \to & H^2(M_3) & \to & H^3(M_4, M_3) & \to & H^3(M_4) \to 0 \\ \| & & \| & & & & \| & & \| \\ \mathbb{Z}^{b_2} \oplus T_2 & & \mathbb{Z}^{b_2} \oplus T_1 & & & & T_1 & & T_2 \end{array} \tag{27}$$

where T_1 and T_2 are torsion groups. Since $T_2 \to T_1$ is injective, one can introduce $t = |T_1|/|T_2|$. Then,

$$|H_1(M_3; \mathbb{Z})| \; = \; t^2 |\det Q| \tag{28}$$

In particular, when both torsion groups T_1 and T_2 are trivial, we simply have a short exact sequence

$$0 \longrightarrow \Gamma \xrightarrow{Q} \Gamma^* \longrightarrow H^2(M_3) \longrightarrow 0 \tag{29}$$

so that $H_1(M_3) \cong H^2(M_3)$ is isomorphic to $\mathbb{Z}^{b_2}/Q(\mathbb{Z}^{b_2})$, generated by the meridians μ_i of the link components K_i, modulo relations imposed by the intersection form Q of the 4-manifold (13):

$$H_1(M_3;\mathbb{Z}) \ = \ \mathbb{Z}[\mu_1,\dots,\mu_n]/\mathrm{im}Q \tag{30}$$

It follows that, in the case of $G = U(1)$ (i.e., a single fivebrane), the representation variety (24) is parametrized by the eigenvalues $x_i \in \mathbb{C}^*$ of the $G_{\mathbb{C}}$-valued holonomies along the 1-cycles μ_i, subject to the relations in (30):

$$\prod_{i=1}^{n} x_i^{Q_{ij}} \ = \ 1 \qquad \forall j = 1,\dots,n \tag{31}$$

There is a similar description of $\mathcal{V}_{T[M_3;G]}$ for non-abelian groups as well [Guk05]. One important consequence of this calculation is that $H_1(M_3;\mathbb{Z})$ is finite and, therefore, the 3d $\mathcal{N} = 2$ theory $T[M_3]$ has finitely many vacua if and only if all eigenvalues of the intersection form Q_{M_4} are non-zero. If Q has zero eigenvalues, then $H_1(M_3;\mathbb{Z})$ contains free factors. This happens, for example, for knots with zero framing coefficients, $a = 0$. Every such Kirby diagram leads to a boundary 3-manifold M_3, whose first homology group is generated by the meridian μ of the knot K with no relations. This clarifies, for instance, why the boundary of a 4-manifold shown in Fig. 3 has $H_1(M_3;\mathbb{Z}) \cong \mathbb{Z}^3$.

If M_4 is a sphere plumbing represented by a plumbing tree Υ, then the eigenvalues of Q can be obtained using a version of the Gauss algorithm that consists of the following two simple steps (see, e.g., [Sav02]):

1. Pick any vertex in Υ and orient all edges toward it. Since Υ is a tree, this is always possible.
2. Recursively applying the rules in Fig. 4 remove the edges, replacing the integer weights a_i (= framing coefficients of the original Kirby diagram) by rational weights.

In the end of this process, when there are no more edges left, the rational weights r_i are precisely the eigenvalues of the intersection form Q and

$$\det(Q) \ = \ \prod_i r_i \tag{32a}$$

$$\mathrm{sign}(Q) \ = \ \#\{i|r_i > 0\} - \#\{i|r_i < 0\} \tag{32b}$$

For example, applying this algorithm to the plumbing tree in Fig. 5 we get

$$\det(Q) \ = \ \left(b + \sum_{i=1}^{k} \frac{q_i}{p_i}\right) \cdot \prod_{i=1}^{k} p_i \tag{33}$$

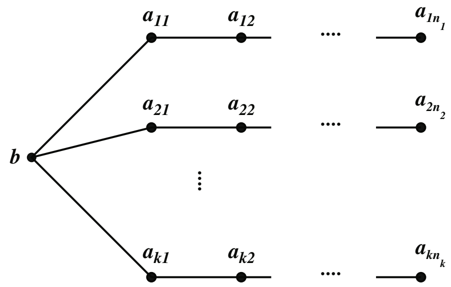


Fig. 5 Plumbing tree of a 4-manifold bounded by a Seifert fibration. We assume $b \le -1$ and $a_{ij} \le -2$

where $-\frac{p_i}{q_i} = [a_{i1}, \dots, a_{in_i}]$ are given by the continued fractions

$$-\frac{p_i}{q_i} = a_{i1} - \cfrac{1}{a_{i2} - \cfrac{1}{\ddots - \cfrac{1}{a_{in_i}}}} \tag{34}$$

The boundary 3-manifold in this case is the Seifert fibered homology 3-sphere $M_3(b; (p_1, q_1), \dots, (p_k, q_k))$ with singular fibers of orders $p_i \ge 1$. It is known that any Seifert fibred rational homology sphere bounds at least one definite form. In our applications here, we are mostly interested in the choice of orientation, such that a Seifert manifold M_3 bounds a plumbed 4-manifold with negative definite intersection form. Then, M_3 is the link of a complex surface singularity.

2.4 *Quiver Chern–Simons Theory*

We already mentioned a striking similarity between plumbing graphs and quivers. The latter are often used to communicate quickly and conveniently the field content of gauge theories, in a way that each node of the quiver diagram represents a simple Lie group and every edge corresponds to a bifundamental matter. Here, we take this hint a little bit more seriously and, with a slight modification of the standard rules, associate a 3d $\mathcal{N} = 2$ gauge theory to a plumbing graph Υ, which will turn out to be an example of the sought-after theory $T[M_3]$.

Much as in the familiar quiver gauge theories, to every vertex of Υ we are going to associate a gauge group factor. Usually, the integer label of the vertex represents the rank. In our present example, however, we assign to each vertex a gauge group $U(1)$ with pure $\mathcal{N} = 2$ Chern–Simons action at level k determined by the integer weight (= the framing coefficient) of that vertex:

$$\begin{aligned} S &= \frac{k}{4\pi} \int d^3x d^4\theta \; V\Sigma \\ &= \frac{k}{4\pi} \int (A \wedge dA - \overline{\lambda}\lambda + 2D\sigma) \end{aligned} \tag{35}$$

Here, $V = (A_\mu, \lambda, \sigma, D)$ is the three-dimensional $\mathcal{N} = 2$ vector superfield and $\Sigma = \overline{D}^\alpha D_\alpha V$ is the field strength superfield.

Similarly, to every edge of Υ that connects a vertex "i" with a vertex "j" we associate 3d $\mathcal{N} = 2$ Chern–Simons coupling between the corresponding vector superfields V_i and V_j:

$$S = \frac{1}{2\pi} \int d^3x d^4\theta \; V_i \Sigma_j \tag{36}$$

Both of these basic building blocks can be combined together with the help of the symmetric bilinear form (20). As a result, to a plumbing graph Υ we associate the following 3d $\mathcal{N} = 2$ theory:

$$T[M_3; U(1)] \;=\; \begin{cases} U(1)^n \text{ quiver Chern–Simons theory with Lagrangian} \\ \mathcal{L} = \sum_{i,j=1}^{n} \int d^4\theta \, \frac{Q_{ij}}{4\pi} \, V_i \Sigma_j \;=\; \frac{1}{4\pi} \int Q(A, dA) + \ldots \end{cases} \tag{37}$$

where $n = \text{rank}(Q)$ and the ellipses represent $\mathcal{N} = 2$ supersymmetric completion of the bosonic Chern–Simons action. Note, since the gauge group is abelian, the fermions in the $\mathcal{N} = 2$ supersymmetric completion of this Lagrangian decouple. As for the bosonic part, quantum-mechanically it only depends on the discriminant group of the lattice (Γ, Q),

$$\mathfrak{D} \;=\; H_1(M_3; \mathbb{Z}) \tag{38}$$

and a $\mathbb{Q}/\mathbb{Z}$-valued quadratic form $\mathfrak{q}$ on $\mathfrak{D}$ [KS11].

We claim that the quiver Chern–Simons theory (37) provides a Lagrangian description of the 3d $\mathcal{N} = 2$ theory $T[M_3; U(1)]$ for *any* boundary 3-manifold M_3. Indeed, by a theorem of Rokhlin, every closed oriented 3-manifold M_3 bounds a 4-manifold of the form (13) and can be realized as an integral surgery on some link in S^3. Denoting by Q the intersection form (resp. the linking matrix) of the corresponding 4-manifold (resp. its Kirby diagram), we propose 3d $\mathcal{N} = 2$ theory (37) with Chern–Simons coefficients Q_{ij} to be a Lagrangian description of the boundary theory $T[M_3; U(1)]$.

To justify this proposal, we note that supersymmetric vacua of the theory (37) on $S^1 \times \mathbb{R}^2$ are in one-to-one correspondence with solutions to (31). Indeed, upon reduction on a circle, each 3d $\mathcal{N} = 2$ vector multiplet becomes a twisted chiral multiplet, whose complex scalar component we denote $\sigma_i = \log x_i$. The Chern–Simons coupling (37) becomes the twisted chiral superpotential, see, e.g., [DGG1, FGP13]:

$$\widetilde{\mathcal{W}} \;=\; \sum_{i,j=1}^{n} \frac{Q_{ij}}{2} \log x_i \cdot \log x_j \tag{39}$$

Extremizing the twisted superpotential with respect to the dynamical fields $\sigma_i = \log x_i$ gives equations for supersymmetric vacua:

$$\exp\left(\frac{\partial\widetilde{\mathcal{W}}}{\partial \log x_i}\right) = 1 \tag{40}$$

which reproduce (31).

2.5 The Lens Space Theory

Of particular importance to the construction of two-dimensional theories $T[M_4]$ are special cases that correspond to 4-manifolds bounded by Lens spaces $L(p,q)$. We remind that the Lens space $L(p,q)$ is defined as the quotient of $S^3 = \{(z_1,z_2) \in \mathbb{C}^2 \mid |z_1|^2 + |z_2|^2 = 1\}$ by a $\mathbb{Z}_p$-action generated by

$$(z_1,z_2) \sim (e^{2\pi i/p}z_1, e^{2\pi iq/p}z_2) \tag{41}$$

We assume p and q to be coprime integers in order to ensure that $\mathbb{Z}_p$-action is free and the quotient is smooth. Two Len's spaces $L(p,q_1)$ and $L(p,q_2)$ are homotopy equivalent if and only if $q_1q_2 \equiv \pm n^2 \mod p$ for some $n \in \mathbb{N}$, and homeomorphic if and only if $q_1 \equiv \pm q_2^{\pm 1} \mod p$. Reversing orientation means $L(p,-q) = -L(p,q)$. Note, supersymmetry (of the cone built on the Lens space) requires $q + 1 \equiv 0 \mod p$.

In the previous discussion we already encountered several examples of 4-manifolds bounded by Lens spaces. These include the disk bundle over S^2 with the Kirby diagram (15) and the linear plumbing on A_{p-1}, which are bounded by $L(p,1)$ and $L(p,-1)$, respectively. In particular, for future reference we write

$$\partial A_p = L(p+1,p) \tag{42}$$

In fact, a more general linear plumbing of oriented circle bundles over spheres with Euler numbers $a_1, a_2, \ldots, a_n$ (see Fig. 1) is bounded by a Lens space $L(p,q)$, such that $[a_1, a_2, \ldots, a_n]$ is a continued fraction expansion for $-\frac{p}{q}$,

$$-\frac{p}{q} = a_1 - \cfrac{1}{a_2 - \cfrac{1}{\ddots - \cfrac{1}{a_n}}} \tag{43}$$

When $p > q > 0$ we may restrict the continued fraction coefficients to be integers $a_i \leq -2$, for all $i = 1, \ldots, n$, so that $L(p,q)$ is the oriented boundary of the negative definite plumbing associated with the string $(a_1, a_2, \ldots, a_n)$. With these orientation conventions, the Lens space $L(p, q)$ is defined by a $(-\frac{p}{q})$-surgery on an unknot in S^3. We also point out that any lens space $L(p, q)$ bounds both positive and negative definite forms Q. (Note, according to the Donaldson's theorem [Don83], the only definite forms that S^3 bounds are the diagonal unimodular forms.)

Next, let us discuss 3d $\mathcal{N} = 2$ theory $T[M_3; G]$ for $M_3 = L(p, q)$ and $G = U(N)$. First, since $H_1(M_3) = \mathbb{Z}_p$ we immediately obtain the number of vacua on $S^1 \times \mathbb{R}^2$, cf. (25):

$$\#\{\text{vacua of } T[L(p,q); U(N)]\} \; = \; \frac{(N+p-1)!}{N!(p-1)!} \tag{44}$$

which, according to (24), is obtained by counting $U(N)$ flat connections on $S^3/\mathbb{Z}_p$. Incidentally, this also equals the number of $SU(p)$ representations at level N, which is crucial for identifying Vafa–Witten partition functions on ALE spaces with WZW characters [Nak94, VW94].

There are several ways to approach the theory $T[L(p, q); U(N)]$, in particular, to give a Lagrangian description, that we illustrate starting with the simple case of $N = 1$ and $q = 1$. For example, one approach is to make use of the Hopf fibration structure on the Lens space $L(p, 1) = S^3/\mathbb{Z}_p$ and to reduce the M-theory setup with a fivebrane on the S^1 fiber. This reduction was very effective, e.g., in analyzing a similar system of fivebranes on Lens spaces with half as much supersymmetry [AV01]. It yields type IIA string theory with a D4-brane wrapped on the base S^2 of the Hopf fibration with $-p$ units of Ramond–Ramond 2-form flux through the S^2. The effective theory on the D4-brane is 3d $\mathcal{N} = 2$ theory with $U(1)$ gauge group and supersymmetric Chern–Simons coupling at level $-p$ induced by the RR 2-form flux, thus, motivating the following proposal:

$$T[L(p,1); U(1)] \; = \; U(1) \text{ SUSY Chern–Simons theory at level } -p \tag{45}$$

To be more precise, this theory as well as quiver Chern–Simons theories (37) labeled by plumbing graphs in addition includes free chiral multiplets, one for each vertex in the plumbing graph. Since in the abelian, $G = U(1)$ case these chiral multiplets decouple and do not affect the counting of $G_{\mathbb{C}}$ flat connections, we tacitly omit them in our present discussion. However, they play an important role and need to be included in the case of $G = U(N)$.

Another approach, that also leads to (45), is based on the Heegaard splitting of M_3. Indeed, as we already mentioned earlier, $L(p, q)$ is a Dehn surgery on the unknot in S^3 with the coefficient $-\frac{p}{q}$. It means that $M_3 = L(p, q)$ can be glued from two copies of the solid torus, $S^1 \times D^2$, whose boundaries are identified via non-trivial map $\phi : T^2 \to T^2$. The latter is determined by its action on homology $H_1(T^2; \mathbb{Z}) \cong \mathbb{Z} \oplus \mathbb{Z}$ which, as usual, we represent by a 2×2 matrix

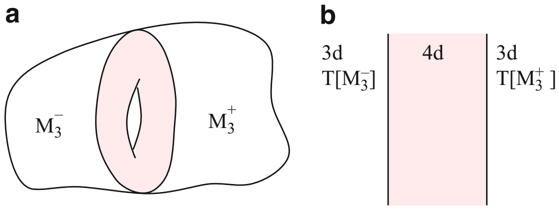


Fig. 6 (**a**) A genus-1 Heegaard splitting of a 3-manifold M_3 becomes a 4d $\mathcal{N} = 4$ super-Yang-Mills theory (**b**) coupled to three-dimensional $\mathcal{N} = 2$ theories $T[M_3^-]$ and $T[M_3^+]$ at the boundary

$$\phi = \begin{pmatrix} p & r \\ q & s \end{pmatrix} \tag{46}$$

with $ps - qr = 1$. If $(-\frac{p}{q}) = [a_1, a_2, \dots, a_n]$ is given by the continued fraction expansion (43), we can explicitly write

$$\begin{pmatrix} p & r \\ q & s \end{pmatrix} = \begin{pmatrix} -a_1 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -a_2 & -1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} -a_n & -1 \\ 1 & 0 \end{pmatrix} \tag{47}$$

This genus-1 Heegaard decomposition has a simple translation to physics, illustrated in Fig. 6. Again, let us first consider the simple case with $N = 1$ and $q = 1$. Then, the 6d (0, 2) theory on T^2 gives 4d $\mathcal{N} = 4$ supersymmetric Maxwell theory, in which the $SL(2, \mathbb{Z})$ action (46) on a torus is realized as the electric–magnetic duality transformation. On the other hand, each copy of the solid torus defines a "Lagrangian" boundary condition that imposes Dirichlet boundary condition on half of the $\mathcal{N} = 4$ vector multiplet and Neumann boundary condition on the other half. Hence, the combined system that corresponds to the Heegaard splitting of $L(p, 1)$ is 4d $\mathcal{N} = 4$ Maxwell theory on the interval with two Lagrangian boundary conditions that are related by an S-duality transformation $\phi = \left(\begin{smallmatrix} p & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and altogether preserve $\mathcal{N} = 2$ supersymmetry in three non-compact dimensions.

Following the standard techniques [HW97, GW09], this theory can be realized on the world-volume of a D3-brane stretched between two fivebranes, which impose suitable boundary conditions at the two ends of the interval. If both boundary conditions were the same, we could take both fivebranes to be NS5-branes. However, since in this brane approach the S-duality of $\mathcal{N} = 4$ gauge theory is realized as S-duality of type IIB string theory, it means that the two fivebranes on which D3-brane ends are related by a transformation (46). In particular, if we choose one of the fivebranes to be NS5, then the second fivebrane must be a (p, q) fivebrane, with D5-brane charge p and NS5-brane charge q, as shown in Fig. 7. In the present case, $q = 1$ and the effective theory on the D3-brane stretched between NS5-brane and a 5-brane of type $(p, 1)$ is indeed $\mathcal{N} = 2$ abelian Chern–Simons theory (35) at level $-p$, in agreement with (45).

This approach based on Heegaard splitting and the brane construction suggests that $T[L(p, q); U(1)]$ associated with a more general gluing automorphism (46) should be a 3d $\mathcal{N} = 2$ theory on the D3-brane stretched between NS5-brane and a 5-brane of type (p, q). This theory on the D3-brane, shown in Fig. 7, indeed

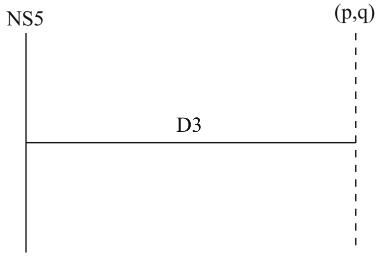


Fig. 7 The effective 3d $\mathcal{N}=2$ theory on a D3-brane stretched between NS5-brane and a 5-brane of type (p,q) is a Chern–Simons theory at level $k=-\frac{p}{q}$. We describe it as a "quiver Chern–Simons theory" with *integer* levels a_i given by the continued fraction $-\frac{p}{q}=[a_1,\ldots,a_n]$

has the effective Chern–Simons coupling at level $-\frac{p}{q}$ [KOO99, BHKK99, Oht99]. However, a better way to think about this $\mathcal{N}=2$ theory—that avoids using fractional Chern–Simons levels and that we take as a proper Lagrangian formulation of $T[L(p,q);U(1)]$—is based on writing the general $SL(2,\mathbb{Z})$ element (46) as a word in standard S and T generators that obey $S^4=(ST)^3=\mathrm{id}$,

$$\phi \;=\; ST^{a_1}\,ST^{a_2}\,\cdots\,ST^{a_n} \tag{48}$$

and implementing it as a sequence of operations on the 3d $\mathcal{N}=2$ abelian gauge theory *a la* [Wit03]. Specifically, the T element of $SL(2,\mathbb{Z})$ acts by adding a level-1 Chern–Simons term,

$$T:\quad \Delta\mathcal{L}=\frac{1}{4\pi}\int d^4\theta\; V\Sigma=\frac{1}{4\pi}A\wedge dA+\cdots \tag{49}$$

while the S transformation introduces a new $U(1)$ gauge (super)field $\widetilde{A}$ coupled to the "old" gauge (super)field A via Chern–Simons term

$$S:\quad \Delta\mathcal{L}=\frac{1}{2\pi}\int d^4\theta\; \widetilde{V}\Sigma=\frac{1}{2\pi}\widetilde{A}\wedge dA+\cdots \tag{50}$$

Equivalently, the new vector superfield containing $\widetilde{A}$ couples to the "topological" current $*F=*dA$ carried by the magnetic charges for A.

Using this $SL(2,\mathbb{Z})$ action on abelian theories in three dimensions, we propose the following candidate for the generalization of the Lens space theory (45) to $|q|\geq 1$:

$$T[L(p,q);U(1)] \;=\; U(1)^n \text{ theory with Chern–Simons coefficients } Q_{ij} \tag{51}$$

where the matrix Q is given by (18) and $-\frac{p}{q}=[a_1,\ldots,a_n]$ is the continued fraction expansion (43). Note, the matrix of Chern–Simons coefficients in this Lens space theory can be conveniently represented by a quiver diagram identical to the plumbing graph in Fig. 1. The proposal (51) for the Lens space theory is, in fact, a special case of (37) and can be justified in the same way, by comparing the critical points of the twisted superpotential (39) with solutions to (31).

Both methods that we used to derive the basic 3d $\mathcal{N} = 2$ Lens space theory (45) suggest a natural generalization to $G = U(N)$:

$$T[L(p,1); U(N)] \;=\; \begin{cases} U(N) \text{ SUSY Chern–Simons theory at level } -p \\ \text{with a chiral multiplet in the adjoint representation} \end{cases} \tag{52}$$

which corresponds to replacing a single D3-brane in the brane construction on Fig. 7 by a stack of N D3-branes. Indeed, the Witten index of $\mathcal{N} = 2$ Chern–Simons theory with gauge group $SU(N)$ and level p (with or without super-Yang-Mills term) is equal to the number of level p representations of affine $SU(N)$, see [Wit99] and also [BHKK99, Oht99, Smi10]:

$$\mathcal{I}_{SU(N)_p} \;=\; \frac{(N+p-1)!}{(N-1)!p!} \tag{53}$$

After multiplying by $\frac{p}{N}$ to pass from the gauge group $SU(N)$ to $U(N) = \frac{U(1)\times SU(N)}{\mathbb{Z}_N}$ we get the number of $SU(p)_N$ representations (44), which matches the number of $U(N)$ flat connections on the Lens space $L(p,1)$. Note that the role of the level and the rank are interchanged compared to what one might naturally expect. An alternative UV Lagrangian for the theory (52), that makes contact with the cohomology of the Grassmannian [Wit93, KW13], is a $\mathcal{N} = 2$ $U(N)$ Chern–Simons action at level $-\frac{p}{2}$ coupled to a chiral multiplet in the adjoint representation and p chiral multiplets in the anti-fundamental representation. This theory was studied in detail in [GP15], where further connections to integrable systems and quantum equivariant K-theory of vortex moduli spaces were found.

2.6 3d $\mathcal{N} = 2$ Theory $T[M_3; G]$ for General M_3 and G

Now it is clear how to tackle the general case of N fivebranes on a 4-manifold M_4 with boundary $M_3 = \partial M_4$. This setup leads to a 2d $\mathcal{N} = (0,2)$ theory $T[M_4; G]$ on the boundary of the half-space coupled to a 3d $\mathcal{N} = 2$ theory $T[M_3; G]$ in the bulk, with the group G of rank N and Cartan type A, D, or E.

For a general class of 4-manifolds (13) considered here, the boundary 3-manifold is an integral surgery on a link K in S^3. As usual, we denote the link components K_i, $i = 1, \dots, n$. Therefore, the corresponding theory $T[M_3]$ can be built by "gluing" the 3d $\mathcal{N} = 2$ theory $T[S^3 \setminus K]$ associated with the link complement with n copies of the 3d $\mathcal{N} = 2$ theory $T[S^1 \times D^2]$ associated with the solid torus:

$$T[M_3] \;=\; T[S^3 \setminus K] \otimes \underbrace{\Big(\phi_{a_1} \circ T[S^1 \times D^2]\Big) \otimes \ldots \otimes \Big(\phi_{a_n} \circ T[S^1 \times D^2]\Big)}_{n \text{ copies}} \tag{54}$$

As pointed out in the footnote 8, it is important that the theory $T[S^3 \setminus K]$ accounts for *all* flat $G_{\mathbb{C}}$ connections on the link complement, including the abelian ones. Such theories are known for $G_{\mathbb{C}} = SL(2,\mathbb{C})$ and for many simple knots and links [NRXS12, FGSS], in fact, even in a more "refined" form that knows about categorification and necessarily incorporates all branches of flat connections. For $G_{\mathbb{C}}$ of higher rank, it would be interesting to work out such $T[S^3 \setminus K]$ following [DGG13]. In particular, the results of [DGG13] elucidate one virtue of 3d $\mathcal{N} = 2$ theories $T[M_3; G]$: they always seem to admit a UV description with only $U(1)$ gauge fields (but possibly complicated matter content and interactions). This will be especially important to us in Sect. 4: in order to identify a 2d $(0,2)$ theory $T[M_4]$ associated with a 4-manifold M_4 bounded by M_3 we only need to understand boundary conditions of abelian 3d $\mathcal{N} = 2$ theories.

The second basic ingredient in (54) is the theory $T[S^1 \times D^2]$ associated with the solid torus. This theory is very simple for any $N \geq 1$ and corresponds to the Dirichlet (D5-brane) boundary condition of $\mathcal{N} = 4$ super-Yang-Mills theory, cf., Fig. 6. To be more precise, if we denote by $\mathbb{T} \subset G$ the maximal torus of G, then $G_{\mathbb{C}}$ flat connections on $T^2 = \partial\big(S^1 \times D^2\big)$ are parametrized by two $\mathbb{T}_{\mathbb{C}}$-valued holonomies, modulo the Weyl group W of G,

$$(x, y) \;\in\; (\mathbb{T}_{\mathbb{C}} \times \mathbb{T}_{\mathbb{C}})\,/W \tag{55}$$

Only a middle dimensional subvariety in this space corresponds to $G_{\mathbb{C}}$ flat connections that can be extended to the solid torus $S^1 \times D^2$. Namely, since one of the cycles of T^2 (the meridian of K_i) is contractible in $N(K_i) \cong S^1 \times D^2$, the $G_{\mathbb{C}}$ holonomy on that cycle must be trivial, i.e.,

$$\mathcal{V}_{T[S^1 \times D^2]} \;=\; \left\{ (x_i, y_i) \in \frac{\mathbb{T}_{\mathbb{C}} \times \mathbb{T}_{\mathbb{C}}}{W} \;\middle|\; x_i = 1 \right\} \tag{56}$$

The $SL(2,\mathbb{Z})$ transformation ϕ_{a_i} gives a slightly more interesting theory $\phi_{a_i} \circ T[S^1 \times D^2]$, whose space of supersymmetric vacua (24) is simply an $SL(2,\mathbb{Z})$ transform of (56):

$$\mathcal{V}_{\phi_{a_i} \circ T[S^1 \times D^2]} \;=\; \left\{ (x_i, y_i) \in \frac{\mathbb{T}_{\mathbb{C}} \times \mathbb{T}_{\mathbb{C}}}{W} \;\middle|\; x_i^{a_i} y_i = 1 \right\} \tag{57}$$

See, e.g., [Guk05] for more details on Dehn surgery in the context of complex Chern–Simons theory.

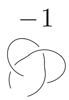
The space of vacua (57) essentially corresponds to $\mathcal{N} = 2$ Chern–Simons theory at level a_i. Therefore, when performing a surgery on K_i, the operation of gluing back $N(K_i) \cong S^1 \times D^2$ with a twist $\phi_{a_i} \in SL(2,\mathbb{Z})$ means gauging the i-th global symmetry of the 3d $\mathcal{N} = 2$ theory $T[S^3 \setminus K]$ and introducing a Chern–Simons term at level a_i. Before this operation, in the theory $T[S^3 \setminus K]$ associated with the link complement, the twisted masses and Fayet–Iliopoulos parameters $(\log x_i, \log y_i)$ are expectation values of real scalars in background vector multiplets that couple to flavor and topological currents, respectively

For instance, when $G_{\mathbb{C}} = SL(2,\mathbb{C})$ and K is a knot (i.e., a link with a single component), the holonomy eigenvalues x and y are both $\mathbb{C}^*$-valued, and the space of vacua $\mathcal{V}_{T[S^3 \setminus K]}$ is the algebraic curve $A_K(x,y) = 0$, the zero locus of the A-polynomial. Therefore, modulo certain technical details, the vacua of the combined theory (54), in this case can be identified with the intersection points of the two algebraic curves, cf. (26):

$$\mathcal{V}_{T[M_3]} \;=\; \{A_K(x,y) = 0\} \cap \{x^a y = 1\} \tag{58}$$

modulo $\mathbb{Z}_2$ action of the $SL(2,\mathbb{C})$ Weyl group $(x,y) \mapsto (x^{-1}, y^{-1})$. Note, both the A-polynomial $A_K(x,y)$ of any knot and the equation $x^a y = 1$ are invariant under this symmetry. In particular, if K is the unknot we have $A(\text{unknot}) = y - 1$ and these two conditions give an $SL(2,\mathbb{C})$ analogue of (31).

As a simple illustration one can consider, say, a negative definite 4-manifold whose Kirby diagram consists of the left-handed trefoil knot $K = \mathbf{3_1}$ with the framing coefficient $a = -1$:



(59)

Using standard tools in Kirby calculus (that we review shortly), it is easy to verify that the boundary of this 4-manifold is the Poincaré homology sphere $\Sigma(2,3,5)$, cf. (21), realized here as a -1 surgery on the trefoil knot in S^3. Therefore, the corresponding theory $T[\Sigma(2,3,5)]$ can be constructed as in (54). The knot complement theory that accounts for all flat connections is well known in this case [FGSS]; in fact, [FGSS] gives two dual descriptions of $T[S^3 \setminus \mathbf{3_1}]$. In this theory, the twisted mass $\log x$ is the vev of the real scalar in background vector multiplet V that couples to the $U(1)_x$ flavor symmetry current. Gauging the flavor symmetry $U(1)_x$ by adding a $\mathcal{N} = 2$ Chern–Simons term for V at level $a = -1$ gives the desired Poincaré sphere theory:

$$\mathcal{L}_{T[\Sigma(2,3,5)]} \;=\; \mathcal{L}_{T[S^3 \setminus \mathbf{3_1}]} - \frac{1}{4\pi} \int d^4\theta \; V\Sigma \tag{60}$$

Upon compactification on S^1, the field $\sigma = \log x$ is complexified and the critical points (40) of the twisted superpotential in the effective 2d $\mathcal{N} = (2,2)$ theory $T[\Sigma(2,3,5)]$,

$$\exp \frac{\partial}{\partial \log x} \left(\widetilde{\mathcal{W}}_{T[S^3 \setminus K]} + \frac{a}{2} (\log x)^2 \right) = 1 \,, \tag{61}$$

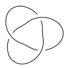
automatically reproduce Eq. (58) for flat $SL(2, \mathbb{C})$ connections.

2.7 Gluing Along a Common Boundary

Given two manifolds M_4^+ and M_4^- which have the same boundary (component) M_3, there is a natural way to build a new 4-manifold labeled by a map $\varphi : M_3 \to M_3$ that provides an identification of the two boundaries:

$$M_4 \;=\; M_4^- \cup_{\varphi} M_4^+ \tag{62}$$

For example, let M_4^- be the negative E_8 plumbing, and let $\overline{M}_4^+$ be the handlebody on the left-handed trefoil knot with the framing coefficient $a = -1$. As we already mentioned earlier, both of these 4-manifolds are bounded by the Poincaré homology sphere $\Sigma(2, 3, 5)$, i.e.,

$E_8 \quad \overset{\partial}{\approx} \quad$ -1  (63)

Therefore, in order to glue these 4-manifolds "back-to-back" as illustrated in Fig. 8, we need to reverse the orientation of one of them, which in the language of Kirby diagrams amounts to replacing all knots with mirror images and flipping the sign of all framing numbers:

$$M_4(K_1^{a_1}, \ldots, K_n^{a_n}) \quad \xrightarrow[\text{reversal}]{\text{orientation}} \quad M_4(\overline{K}_1^{-a_1}, \ldots, \overline{K}_n^{-a_n}) \tag{64}$$

Thus, in our example we need to change the left-handed trefoil knot $K = \mathbf{3_1}$ with framing $a = -1$ to the right-handed trefoil knot $\overline{K}$ with framing coefficient $+1$. The resulting 4-manifold M_4^+ with a single 2-handle that corresponds to this Kirby diagram has boundary $M_3 = \partial M_4^+ = -\partial M_4^-$, so that now it can be glued to $M_4^- = E_8$ plumbing.

Gluing 4-manifolds along a common boundary, as in (62), has a nice physical interpretation. Namely, it corresponds to the following operation on the 2d $\mathcal{N} = (0, 2)$ theories $T[M_4^{\pm}]$ that produces a new theory $T[M_4]$ associated with the resulting 4-manifold $M_4 = M_4^- \cup_{\varphi} M_4^+$. As we already explained in Sect. 2.2, partial topological reduction of the 6d fivebrane theory on a 4-manifold with a boundary M_3 leads to a coupled 2d-3d system of 3d $\mathcal{N} = 2$ theory $T[M_3]$ with a B-type

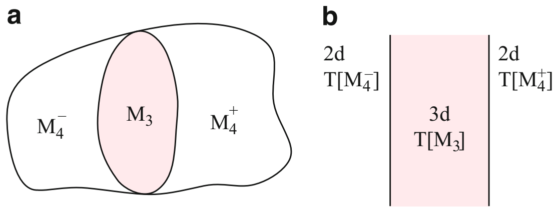


Fig. 8 (**a**) Two 4-manifolds glued along a common boundary $M_3 = \pm\partial M_4^{\pm}$ correspond to (**b**) three-dimensional $\mathcal{N} = 2$ theory $T[M_3]$ on the interval coupled to two-dimensional $\mathcal{N} = (0,2)$ theories $T[M_4^-]$ and $T[M_4^+]$ at the boundaries of the interval

boundary condition determined by the 4-manifold. (If the 4-manifold in question has other boundary components, besides M_3, then the reduction of the 6d fivebrane theory leads to a wall/interface between $T[M_3]$ and other 3d $\mathcal{N} = 2$ theories; this more general possibility will be discussed in the next section.)

In the case at hand, we have two such 4-manifolds, M_4^- and M_4^+, with oppositely oriented boundaries $\partial M_4^{\pm} = \pm M_3$. What this means is that $T[M_4^+]$ defines a B-type boundary condition — with 2d $\mathcal{N} = (0,2)$ supersymmetry on the boundary—in 3d $\mathcal{N} = 2$ theory $T[M_3]$, while $T[M_4^-]$ likewise defines a B-type boundary condition in the theory $T[-M_3]$. Equivalently, $T[-M_3]$ can be viewed as a theory $T[M_3]$ with the reversed parity:

$$T[-M_3] \;=\; P \circ T[M_3] \tag{65}$$

where $P : (x^0, x^1, x^2) \to (x^0, x^1, -x^2)$. This operation, in particular, changes the signs of all Chern–Simons couplings in $T[M_3]$.

Therefore, thanks to (65), we can couple $T[M_4^-]$ and $T[M_4^+]$ to the *same* 3d $\mathcal{N} = 2$ theory $T[M_3]$ considered in space-time $\mathbb{R}^2 \times I$, where I is the interval. In this setup, illustrated in Fig. 8, theories $T[M_4^{\pm}]$ define boundary conditions at the two ends of the interval I. As a result, we get a layer of 3d $\mathcal{N} = 2$ theory $T[M_3]$ on $\mathbb{R}^2 \times I$ sandwiched between $T[M_4^-]$ and $T[M_4^+]$. Since the 3d space-time has only two non-compact directions of $\mathbb{R}^2$, in the infra-red this system flows to a 2d $\mathcal{N} = (0,2)$ theory, which we claim to be $T[M_4]$.

The only element that we need to explain is the map $\varphi : M_3 \to M_3$ that enters the construction (62) of the 4-manifold M_4. If exist, non-trivial self-diffeomorphisms of M_3 correspond to self-equivalences (a.k.a. dualities) of the theory $T[M_3]$. Therefore, a choice of the map $\varphi : M_3 \to M_3$ in (62) means coupling theories $T[M_4^{\pm}]$ to different descriptions/duality frames of the 3d $\mathcal{N} = 2$ theory $T[M_3]$ or, equivalently, inserting a duality wall (determined by φ) into the sandwich of $T[M_4^-]$, $T[M_3]$, and $T[M_4^+]$. Of course, one choice of $\varphi : M_3 \to M_3$ that always exists is the identity map; it corresponds to the most natural coupling of theories $T[M_4^{\pm}]$ to the same description of $T[M_3]$. Since $\varphi : M_3 \to M_3$ can be viewed as a special case of a more general cobordism between two different 3-manifolds that will be discussed in Sect. 2.10, when talking about gluing 4-manifolds we assume that $\varphi = \mathrm{id}$ unless noted otherwise. Then, we only need to know which 4-manifolds have the same boundary.

2.8 3d Kirby Moves

Since our list of operations includes gluing 4-manifolds along their common boundary components, it is important to understand how $M_3(\Upsilon)$ depends on the plumbing graph Υ and which 4-manifolds $M_4(\Upsilon)$ have the same boundary (so that they can be glued together). Not surprisingly, the set of moves that preserve the boundary $M_3(\Upsilon) = \partial M_4(\Upsilon)$ is larger than the set of moves that preserve the 4-manifold $M_4(\Upsilon)$.

Specifically, plumbing graphs Υ_1 and Υ_2 describe the same 3-manifold $M_3(\Upsilon_1) \cong M_3(\Upsilon_2)$ if and only if they can be related by a sequence of "blowing up" or "blowing down" operations shown in Fig. 9, as well as the moves in Fig. 10. The blowing up (resp. blowing down) operations include adding (resp. deleting) a component of Υ that consists of a single vertex with label ± 1. Such blow ups have a simple geometric interpretation as boundary connected sum operations with very simple 4-manifolds $\mathbb{C}\mathbf{P}^2 \setminus \{\mathrm{pt}\}$ and $\overline{\mathbb{C}\mathbf{P}}^2 \setminus \{\mathrm{pt}\}$, both of which have S^3 as a boundary and, therefore, only change M_4 but not $M_3 = \partial M_4$. As will be discussed shortly, this also has a simple physical counterpart in physics of 3d $\mathcal{N} = 2$ theory $T[M_3]$, where the blowup operation adds a decoupled "trivial' ' $\mathcal{N} = 2$ Chern–Simons term (52) at level ± 1, which carries only boundary degrees of freedom and has a single vacuum, cf. (44). For this reason, blowing up and blowing down does not change $T[M_3; G]$ and only changes $T[M_4; G]$ by free Fermi multiplets, for abelian as well as non-abelian G.

Applying these moves inductively, it is easy to derive a useful set of rules illustrated in Fig. 11 that, for purposes of describing the boundary of M_4, allow to collapse linear chains of sphere plumbings with arbitrary framing coefficients a_i via continued fractions

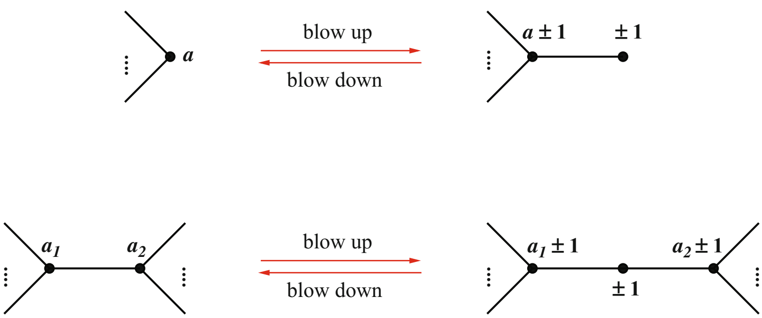


Fig. 9 Blowing up and blowing down does not change the boundary $M_3 = \partial M_4$

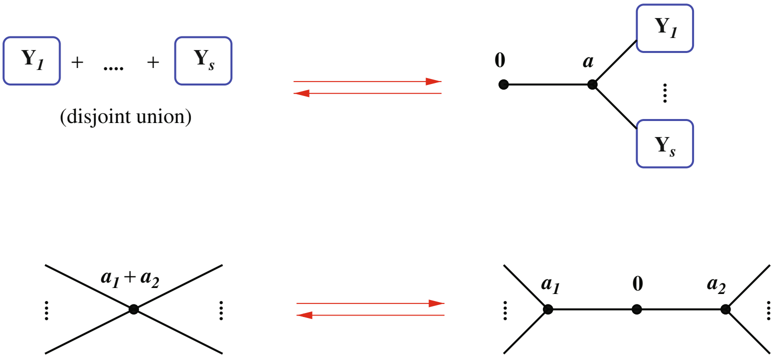


Fig. 10 "3d Kirby moves" that do not change $M_3 = \partial M_4$

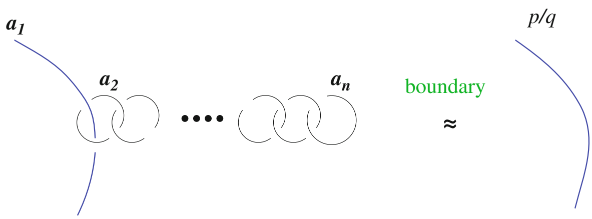


Fig. 11 Boundary diffeomorphisms relating integral surgery and Dehn surgery

$$\frac{p}{q} = a_1 - \cfrac{1}{a_2 - \cfrac{1}{\ddots - \cfrac{1}{a_n}}} \tag{66}$$

To illustrate how this works, let us demonstrate that the A_{n-1} plumbing, as in Fig. 1, with $a_i = -2$ can be glued to a disc bundle with Euler number $-n$ over S^2 to produce a smooth 4-manifold $(\overline{\mathbb{C}\mathbf{P}}^2)^{\# n}$. In particular, we need to show that these two 4-manifolds we are gluing naturally have the same boundary with opposite orientation. This is a simple exercise in Kirby calculus.

Starting with the A_{n-1} linear plumbing, we can take advantage of the fact that ± 1 vertices can be added for free and consider instead

$$\overset{+1}{\bullet} \qquad \overset{-2}{\bullet} — \overset{-2}{\bullet} — \overset{-2}{\bullet} — \quad \cdots \quad — \overset{-2}{\bullet} \tag{67}$$

Clearly, this operation (of blowing up) changes the 4-manifold, but not the boundary M_3. Now, we slide the new $+1$ handle over the -2 handle. According to (14), this preserves the framing $+1$ of the new handle and changes the framing of the -2 handle to $-2 + 1 = -1$ (since they were originally unlinked), resulting in

$$\overset{+1}{\bullet} — \overset{-1}{\bullet} — \overset{-2}{\bullet} — \overset{-2}{\bullet} — \quad \cdots \quad — \overset{-2}{\bullet} \tag{68}$$

Note, this plumbing graph with n vertices is a result of applying the first move in Fig. 9 to the A_{n-1} linear plumbing, which we have explained "in slow motion." Since we now have a vertex with weight -1, we can apply the second move in Fig. 9 to remove this vertex at the cost of increasing the weights of the two adjacent vertices by $+1$, which gives

$$\overset{+2}{\bullet} — \overset{-1}{\bullet} — \overset{-2}{\bullet} — \quad \cdots \quad — \overset{-2}{\bullet} \tag{69}$$

This last step made the plumbing graph shorter, of length $n - 1$, and there is a new vertex with weight $-2 + 1 = -1$ on which we can apply the blow down again. Doing so will change the weight of the leftmost vertex from $+2$ to $+3$ and after $n - 3$ more steps we end up with a plumbing graph

$$\overset{n-1}{\bullet} — \overset{-1}{\bullet} \tag{70}$$

Applying the first move in Fig. 9 we finally get the desired relation

$$A_{n-1} \quad \overset{\partial}{\approx} \quad \overset{+n}{\bullet} \tag{71}$$

Since reversing orientation on the 4-manifold is equivalent (64) to replacing all knots with mirror images and flipping the sign of all framing numbers, this shows that A_{n-1} linear plumbing has the same Lens space boundary as the disc bundle with Euler number $-n$ over S^2, but with opposite orientation. In particular, it follows that these 4-manifolds with boundary can be glued along their common boundary in a natural way. (No additional orientation reversal or other operation is needed.)

Following these arguments, it is easy to show a more general version of the first move in Fig. 9 called *slam-dunk*:

$$\overset{p/q}{\bullet} — \overset{a}{\bullet} — \quad \cdots \qquad \overset{\partial}{\approx} \qquad \overset{a-\frac{q}{p}}{\bullet} — \quad \cdots \tag{72}$$

which, of course, is just a special case of the boundary diffeomorphism in Fig. 11. Another useful rule in 3d Kirby calculus that can be deduced by the same argument allows to collapse a (sub)chain of (-2)'s:

$$— \overset{a}{\bullet} — \underbrace{\overset{-2}{\bullet} — \quad \cdots \quad — \overset{-2}{\bullet}}_{n \text{ times}} — \overset{b}{\bullet} — \qquad \overset{\partial}{\approx} — \overset{a+1}{\bullet} — \overset{n+1}{\bullet} — \overset{b+1}{\bullet} —$$

which is a generalization of (71).

2.9 Physical Interpretation of 3d Kirby Moves

All these moves that preserve the boundary 3-manifold $M_3(\Upsilon) = \partial M_4(\Upsilon)$ have an elegant and simple interpretation as equivalences (dualities) of the corresponding 3d $\mathcal{N} = 2$ theory $T[M_3(\Upsilon); U(N)]$. Let us illustrate this in the basic case of $N = 1$, i.e., a single fivebrane. Then, as we explained in Sect. 2.2, all theories $T[M_3(\Upsilon); U(1)]$ admit a description as supersymmetric Chern–Simons theories, and 3d Kirby moves are precisely the equivalence relations on the matrix of Chern–Simons coefficients in the quantum theory.

Indeed, the simplest version of blowing up (resp. blowing down) operation that adds (resp. removes) an isolated vertex with label ± 1 in the theory $T[M_3(\Upsilon); U(1)]$ correspond to changing the matrix of Chern–Simons coefficients

$$Q \;\to\; Q \oplus \langle \pm 1 \rangle \tag{73}$$

that is, adds (resp. removes) a $U(1)$ vector multiplet V with the Lagrangian

$$\Delta\mathcal{L} \;=\; \pm\frac{1}{4\pi}\int d^4\theta\; V\Sigma \;=\; \pm\frac{1}{4\pi}A \wedge dA + \cdots \tag{74}$$

A theory defined by this Lagrangian is trivial. In particular, it has one-dimensional Hilbert space. Therefore, tensor products with copies of this trivial theory are indeed equivalences of $T[M_3(\Upsilon); U(1)]$. The same is true in the non-abelian case as well, where blowups change $T[M_3; G]$ by "trivial" Chern–Simons terms at level ± 1 that carry only boundary degrees of freedom (and, therefore, only affect the physics of the 2d boundary theory $T[M_4; G]$, but not the 3d bulk theory $T[M_3; G]$).

Similarly, we can consider blowing up and blowing down operations shown in Fig. 9. If in the plumbing graph Υ a vertex with label ± 1 is only linked by one edge to another vertex with label $a \pm 1$, it means that the Lagrangian of the 3d $\mathcal{N} = 2$ theory $T[M_3(\Upsilon); U(1)]$ has the following terms:

$$\mathcal{L} \;=\; \frac{1}{4\pi}\int d^4\theta\,\left(\pm V\Sigma + 2\widetilde{V}\Sigma + (a \pm 1)\widetilde{V}\widetilde{\Sigma} + \cdots\right) \tag{75}$$

where ellipses stand for terms that do not involve the vector superfield V or its field strength Σ. Since the action is Gaussian in V, we can integrate it out by solving the equations of motion $\pm V + \widetilde{V} = 0$. The resulting Lagrangian is

$$\mathcal{L}' \;=\; \frac{1}{4\pi}\int d^4\theta\,\left(\pm\widetilde{V}\widetilde{\Sigma} \mp 2\widetilde{V}\widetilde{\Sigma} + (a \pm 1)\widetilde{V}\widetilde{\Sigma} + \cdots\right) = \frac{1}{4\pi}\int d^4\theta\,\left(a\widetilde{V}\widetilde{\Sigma} + \cdots\right) \tag{76}$$

This gives a physics realization of the blowing up and blowing down operations in the top part of Fig. 9. We can easily generalize it to that in the lower part of Fig. 9. Starting with the right side of the relation, the terms in the Lagrangian which involve the superfield V at Chern–Simons level ± 1 look like

$$\mathcal{L} = \frac{1}{4\pi}\int d^4\theta\,(\pm V\Sigma + 2V_1\Sigma + (a_1 \pm 1)V_1\Sigma_1 + 2V_2\Sigma + (a_2 \pm 1)V_2\Sigma_2 + \cdots) \tag{77}$$

Integrating out V yields $\pm V + V_1 + V_2 = 0$ and the effective Lagrangian

$$\mathcal{L}' = \frac{1}{4\pi}\int d^4\theta\,(a_1V_1\Sigma_1 \mp 2V_1\Sigma_2 + a_2V_2\Sigma_2 + \cdots) \tag{78}$$

which, as expected, describes the left side of the relation in the lower part of Fig. 9. From this physical interpretation of the blowing up and blowing down operations in the $N = 1$ case one can draw a more general lesson: the reason that 2-handles with framing coefficients $a = \pm 1$ are "nice" corresponds to the fact that 3d $\mathcal{N} = 2$ theory $T\big[M_3\big(\overset{\pm 1}{\bullet}\big)\big]$ is trivial.

The physical interpretation of 3d Kirby moves in Fig. 10 is even simpler: 2-handles with framing coefficients $a_i = 0$ correspond to superfields in 3d theory $T[M_3(\Upsilon)]$ that serve as Lagrange multipliers. Again, let us explain this in the basic case of a single fivebrane ($N = 1$). Let us consider the first move in Fig. 10 and, as in the previous discussion, denote by V the $U(1)$ vector superfield associated with a 2-handle (vertex) with framing label 0. Then, the relevant terms in the Lagrangian of the theory $T[M_3(\Upsilon); U(1)]$ associated with the right part of the diagram are

$$\mathcal{L} = \frac{1}{4\pi}\int d^4\theta\,\big(2V\widetilde{\Sigma} + a\widetilde{V}\widetilde{\Sigma} + \cdots\big) \tag{79}$$

Note, there is no Chern–Simons term for V itself, and it indeed plays the role of the Lagrange multiplier for the condition $\widetilde{\Sigma} = 0$. Therefore, integrating out V makes $\widetilde{V}$ pure gauge and removes all Chern–Simons couplings involving $\widetilde{V}$. The resulting quiver Chern–Simons theory is precisely the one associated with the left diagram in the upper part of Fig. 10.

Now, let us consider the second move in Fig. 10, again starting from the right-hand side. The relevant part of the Lagrangian for $T[M_3(\Upsilon); U(1)]$ looks like

$$\mathcal{L} = \frac{1}{4\pi}\int d^4\theta\,(2V\Sigma_1 + a_1V_1\Sigma_1 + 2V\Sigma_2 + a_2V_2\Sigma_2 + \cdots) \tag{80}$$

where the dependence on V is again only linear. Hence, integrating it out makes the "diagonal" combination $V_1 + V_2$ pure gauge, and for $V' = V_1 = -V_2$ we get

$$\mathcal{L}' = \frac{1}{4\pi}\int d^4\theta\,\big((a_1 + a_2)V'\Sigma' + \cdots\big) \tag{81}$$

which is precisely the Lagrangian of the quiver Chern–Simons theory associated with the plumbing graph in the lower left corner of Fig. 10.

Finally, since all other boundary diffeomorphisms in 3d Kirby calculus follow from these basic moves, it should not be surprising that the manipulation in Fig. 11 as well as the slam-dunk move (72) also admit an elegant physical interpretation.

However, for completeness, and to practice a little more with the dictionary between 3d Kirby calculus and equivalences of 3d $\mathcal{N} = 2$ theories, we present the details here. Based on the experience with the basic moves, the reader might have (correctly) guessed that both the boundary diffeomorphism in Fig. 11 and the slam-dunk move (72) correspond to integrating out vector multiplets.

Specifically, for the plumbing graph on the left side of (72) the relevant terms in the Lagrangian of the theory $T[M_3(\Upsilon); U(1)]$ look like

$$\mathcal{L} = \frac{1}{4\pi} \int d^4\theta \left(\frac{p}{q} V\Sigma + 2\widetilde{V}\Sigma + a\widetilde{V}\widetilde{\Sigma} + \cdots \right) \tag{82}$$

Since there are no other terms in the Lagrangian of $T[M_3(\Upsilon); U(1)]$ that contain the superfield V or its (super)field strength Σ, we can integrate it out. Replacing V by the solution to the equation $\frac{p}{q}V + \widetilde{V} = 0$ gives the Lagrangian for the remaining fields

$$\mathcal{L} = \frac{1}{4\pi} \int d^4\theta \left(\left(a - \frac{q}{p}\right)\widetilde{V}\widetilde{\Sigma} + \cdots \right) \tag{83}$$

which is an equivalent description of the theory $T[M_3(\Upsilon); U(1)]$, in fact, the one associated with the right-hand side of the slam-dunk move (72). By now it should be clear what is going on. In particular, by iterating this process and integrating in or integrating out $U(1)$ vector superfields, it is easy to show that quiver Chern–Simons theories associated with Kirby diagrams in Fig. 11 are indeed equivalent.

2.10 Cobordisms and Domain Walls

Now, it is straightforward to generalize the discussion in previous sections to 4-manifolds with two (or more) boundary components. The lesson we learned is that each boundary component of M_4 corresponds to a coupling with 3d $\mathcal{N} = 2$ theory labeled by that component.

In general, when a 4-manifold M_4 has one or more boundary components, it is convenient to view it as a (co)bordism from M_3^- to M_3^+, where $M_3^{\pm}$ is allowed to be empty or contain several connected components, see Fig. 12a. If $M_3^- = \emptyset$ (or $M_3^+ = \emptyset$), then the corresponding 3d $\mathcal{N} = 2$ theory $T[M_3^-]$ (resp. $T[M_3^+]$) is trivial. And, when $M_3^{\pm}$ has more than one connected component, the corresponding theory $T[M_3^{\pm}]$ is simply a tensor product of 3d $\mathcal{N} = 2$ theories associated with those components. (In fact, we already encountered similar situations, e.g., in (54), when we discussed *3-manifolds* with several boundary components.)

What kind of 2d theory $T[M_4]$ corresponds to a cobordism from M_3^- to M_3^+? There are several ways to look at it. First, trying to erase any distinction between M_3^+ and M_3^-, we can view any such 4-manifold as a cobordism from $\emptyset$ to $M_3^+ \sqcup -M_3^-$, i.e., as a 4-manifold with boundary $M_3 = M_3^+ \sqcup -M_3^-$, thus reducing the problem to

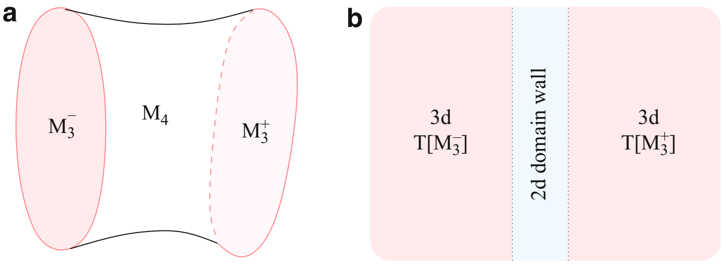


Fig. 12 (**a**) A cobordism between 3-manifolds M_3^- and M_3^+ corresponds to (**b**) a 2d $\mathcal{N} = (0, 2)$ theory $T[M_4]$ on the domain wall (interface) coupled to 3d $\mathcal{N} = 2$ theories $T[M_3^-]$ and $T[M_3^+]$ on both sides

the one already considered. Indeed, using (65), to a 4-manifold M_4 with boundary $M_3^+ \sqcup -M_3^-$ we associate a 3d $\mathcal{N} = 2$ theory $T[M_3^+] \otimes \big(P \circ T[M_3^-]\big)$ on a half-space $\mathbb{R}_+ \times \mathbb{R}^2$ coupled to a boundary theory $T[M_4]$. In turn, this product 3d theory on a half-space is equivalent—via the so-called folding trick [WA94, OA97, BdDO02]—to a 3d theory $T[M_3^+]$ or $T[M_3^-]$ in two regions of the full three-dimensional space $\mathbb{R}^3$, separated by a 2d interface (that in 3d context might be naturally called a "defect wall"). This gives another, perhaps more natural way to think of 2d $\mathcal{N} = (0, 2)$ theory $T[M_4]$ associated with a cobordism from M_3^- to M_3^+, as a theory trapped on the interface separating two 3d $\mathcal{N} = 2$ theories $T[M_3^-]$ or $T[M_3^+]$, as illustrated in Fig. 12.

In order to understand the physics of fivebranes on 4-manifolds, it is often convenient to compactify one more direction, i.e., consider the fivebrane world-volume to be $S^1 \times \mathbb{R} \times M_4$. In the present context, it leads to an effective two-dimensional theory with $\mathcal{N} = (2, 2)$ supersymmetry and a B-type defect[9] labeled by M_4. In fact, we already discussed this reduction on a circle in Sect. 2.2, where it was noted that the effective 2d $\mathcal{N} = (2, 2)$ theory—which, with some abuse of notations, we still denote $T[M_3]$—is characterized by the twisted superpotential $\widetilde{\mathcal{W}}(x_i)$. Therefore, following the standard description of B-type defects in $\mathcal{N} = (2, 2)$ Landau–Ginzburg models [HW04, BR07, BJR08, CR10], one might expect that a defect $T[M_4]$ between two theories $T[M_3^-]$ and $T[M_3^+]$ can be described as a matrix (bi-)factorization of the difference of the corresponding superpotentials

$$\widetilde{W}_{T[M_3^+]}(x_i) - \widetilde{W}_{T[M_3^-]}(y_i) \tag{84}$$

While conceptually quite helpful, this approach is less useful for practical description of the defect walls between $T[M_3^-]$ and $T[M_3^+]$, which we typically achieve by other methods. The reason, in part, is that superpotentials $\widetilde{\mathcal{W}}$ are non-polynomial

[9]The converse is not true since some line defects in 2d come from line operators in 3d.

for most theories $T[M_3]$. We revisit this approach and make additional comments in Sect. 4.

Note, if 2d theories in question were $\mathcal{N} = (2,2)$ sigma-models based on target manifolds $X_{T[M_3^+]}$ and $X_{T[M_3^-]}$, respectively, then B-type defects between them could be similarly represented by correspondences, or (complexes of) coherent sheaves, or sometimes simply by holomorphic submanifolds

$$\Delta \subset X_{T[M_3^+]} \times X_{T[M_3^-]} \tag{85}$$

Much like defect lines in 2d, defect walls in 3d can be classified according to their properties and the symmetries they preserve: topological, conformal, reflective or transmissive, parameter walls, (duality) transformation walls, etc. Various examples of such walls in 3d $\mathcal{N} = 2$ theories were studied in [GGP13]. For instance, parameter walls are labeled by (homotopy types of) paths on the moduli space $\mathcal{V}_{T[M_3]}$ and correspond to (autoequivalence) functors acting on the category of B-type boundary conditions. Transformation walls, on the other hand, in general change 3d $\mathcal{N} = 2$ theory, e.g., by implementing the $SL(2,\mathbb{Z})$ action [Wit03] described in (49)–(50). Topological defects in abelian Chern–Simons theories—which, according to our proposal (37), are relevant to cobordisms between 3-manifolds—have been studied, e.g., in [KS11, KS10, FSV12]. In supersymmetric theories, topological defects are quite special as they are of A-type and B-type at the same time.

The next best thing to topological defects are conformal ones, which in 2d are usually characterized by their reflective or transmissive properties. Extending this terminology to walls in 3d, below we consider two extreme examples, which, much like Neumann and Dirichlet boundary conditions, provide basic ingredients for building mixed types. See Fig. 13a for an illustration of a generic defect wall (neither totally reflective nor fully transmissive).

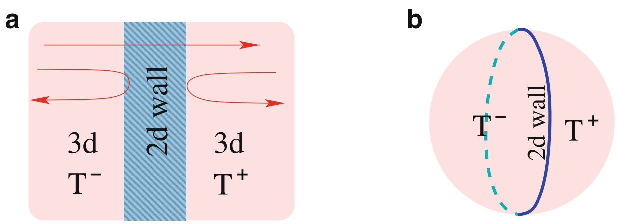


Fig. 13 A generic defect wall between two 3d $\mathcal{N} = 2$ theories (**a**) in flat space-time and (**b**) the corresponding configuration on $S^1 \times S^2$. The index of the latter system is obtained from two copies of the "half-index" $\mathcal{I}_{S^1 \times_q D^\pm}(T^\pm) \simeq Z_{\text{vortex}}(T^\pm)$ convoluted via the index (flavored elliptic genus) of the defect wall supported on $S^1 \times S^1_{\text{eq}}$, where $D^\pm$ is the disk covering right (resp. left) hemisphere of the S^2 and $S^1_{\text{eq}} := \partial D^+ = -\partial D^-$ is the equator of the S^2

2.11 Fully Transmissive Walls

The simplest example of a totally transmissive wall (which is also conformal) is a trivial wall between the theory $T[M_3]$ and itself. It corresponds to the identity cobordism $M_3 \times I$ and in the language of boundary conditions (85) is represented by the "diagonal"

$$\Delta_X \;\subset\; X \times X \tag{86}$$

and similarly for the LG models (84).

In view of (37) and (52), more interesting examples of maximally transmissive defects are walls between $\mathcal{N} = 2$ Chern–Simons theories with gauge groups G and $H \subset G$ that have H-symmetry throughout. Such defects can be constructed by decomposing the Lie algebra

$$\mathfrak{g} = (\mathfrak{g}/\mathfrak{h})^{\perp} \oplus \mathfrak{h}^{\parallel} \tag{87}$$

and imposing Dirichlet type boundary conditions on the coset degrees of freedom and Neumann boundary conditions on degrees of freedom for $H \subset G$. Equivalently, via the level-rank or, in the supersymmetric context, Giveon–Kutasov duality [GK09] equally important are level-changing defect walls in $\mathcal{N} = 2$ Chern–Simons theories. See, e.g., [FSV12] for the study of defect walls with these properties in a purely bosonic theory and [QS02, BM09] for various constructions in closely related WZW models one dimension lower.

2.12 Maximally Reflective Walls

Maximally reflective domain walls between 3d theories $T[M_3^-]$ or $T[M_3^+]$ do not allow these theories to communicate at all. Typical examples of such walls are products of boundary conditions, $\mathcal{B}^-$ and $\mathcal{B}^+$, for $T[M_3^-]$ and $T[M_3^+]$, respectively:

$$T[M_4] \;=\; \mathcal{B}^- \otimes \mathcal{B}^+ \tag{88}$$

In the correspondence between 4-manifolds and 2d $\mathcal{N} = (0, 2)$ theories trapped on the walls, they correspond to disjoint unions $M_4 = M_4^- \sqcup M_4^+$, such that $\partial M_4^{\pm} = M_3^{\pm}$.

2.13 Fusion

Finally, the last general aspect of domain walls labeled by cobordisms that we wish to mention is composition (or, fusion), Illustrated, e.g., in Fig. 15. As we explain in the next section, the importance of this operation is that any 4-manifold of the form (13) and, therefore, any 2d $\mathcal{N} = (0, 2)$ theory associated with it can built—in general, in more than one way—as a sequence of basic fusions. Notice, while colliding general defect walls can be singular, the fusion of B-type walls on $S^1 \times \mathbb{R}^2$ is smooth (since they are compatible with the topological twist along $\mathbb{R}^2$).

2.14 Adding a 2-Handle

We introduced many essential elements of the dictionary (in Table 1) between 4-manifolds and the corresponding 2d theories $T[M_4]$, and illustrated some of them in simple examples. Further aspects of this dictionary and more examples will be given in later sections and future work. One crucial aspect—which, hopefully, is already becoming clear at this stage—is that a basic building block is a 2-handle. Indeed, adding 2-handles one-by-one, we can build *any* 4-manifold of the form (13)! And the corresponding 2d theory $T[M_4]$ can be built in exactly the same way, following a sequence of basic steps, each of which corresponds to adding a new 2-handle.

In this section, we shall look into details of this basic operation and, in particular, explain that adding a new 2-handle at any part of the Kirby diagram can be represented by a cobordism. Then, using the dictionary between cobordisms and walls (interfaces) in 3d, that we already explained in Sect. 2.10, we learn that the operation of adding a 2-handle can be described by a fusion with the corresponding wall, as illustrated in Figs. 14 and 15.

This interpretation of adding 2-handles is very convenient and very powerful, especially for practical ways of building theories $T[M_4]$. For instance, it can be used to turn a small sample of concrete examples into a large factory for producing many new ones. Indeed, suppose one has a good understanding of a (possibly rather small) family of 4-manifolds that can be obtained from one another by adding 2-handles. Then, by extracting[10] the "difference" one gets a key to a much larger class of 4-manifolds and the corresponding theories $T[M_4]$ that can be constructed by composing the basic steps (of adding 2-handles) in a variety of new ways, thus, potentially taking us well outside of the original family. A good starting point for implementing this algorithm and deducing the set of basic cobordisms (resp. the 2d $(0, 2)$ domain wall theories) can be a class of ADE sphere plumbings, as in Figs. 1 and 2, for which the Vafa–Witten partition function is known to be the level N character of the corresponding WZW model [Nak94, VW94]. We pursue this approach in Sect. 3 and identify the corresponding basic operations of adding 2-handles with certain coset models.

Suppose our starting point is a 4-manifold M_4^- with boundary

$$\partial M_4^- = M_3^- \tag{89}$$

[10]Explaining how to do this is precisely the goal of the present section.

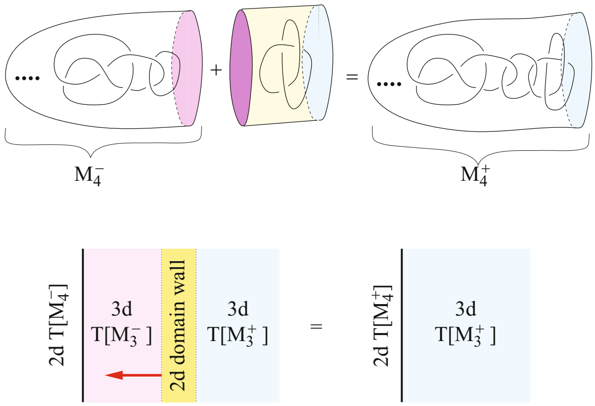


Fig. 14 The operation of attaching a 2-handle to M_4^- can be represented by a cobordism, namely the closure of $M_4^+ \setminus M_4^-$. This operation corresponds to fusing a 2d wall (interface) determined by the cobordism with a boundary theory $T[M_4^-]$ to produce a new boundary theory $T[M_4^+]$. Equivalently, the system on the left—with a domain wall sandwiched between 3d $\mathcal{N} = 2$ theories $T[M_3^-]$ and $T[M_3^+]$—flows in the infra-red to a new boundary condition determined by $T[M_4^+]$

Attaching to it an extra 2-handle we obtain a new 4-manifold M_4^+ with a new boundary

$$\partial M_4^+ = M_3^+ \tag{90}$$

A convenient way to describe this operation—which admits various generalizations and a direct translation into operations on $T[M_4^-]$—is to think of (the closure of) $M_4^+ \setminus M_4^-$ as a (co)bordism, B, from M_3^- to M_3^+. In other words, we can think of M_4^+ as a 4-manifolds obtained by gluing M_4^- to a cobordism B with boundary

$$\partial B = -M_3^- \cup M_3^+ \tag{91}$$

Therefore,

$$M_4^+ \;\; = \;\; M_4^- \cup_\varphi B \tag{92}$$

where $\varphi : M_3 \to M_3$ is assumed to be the identity map, unless noted otherwise.

We have $H_3(M_4^+, B) \cong H_3(M_4^-, M_3^-) \cong H^1(M_4^-)$ by Poincaré duality. The latter is trivial, $H^1(M_4^-) = 0$. Then, comparing the exact sequence for the pair (M_4^+, B) with the exact sequence for the triple (M_4^+, B, M_3^+) we get the following diagram:

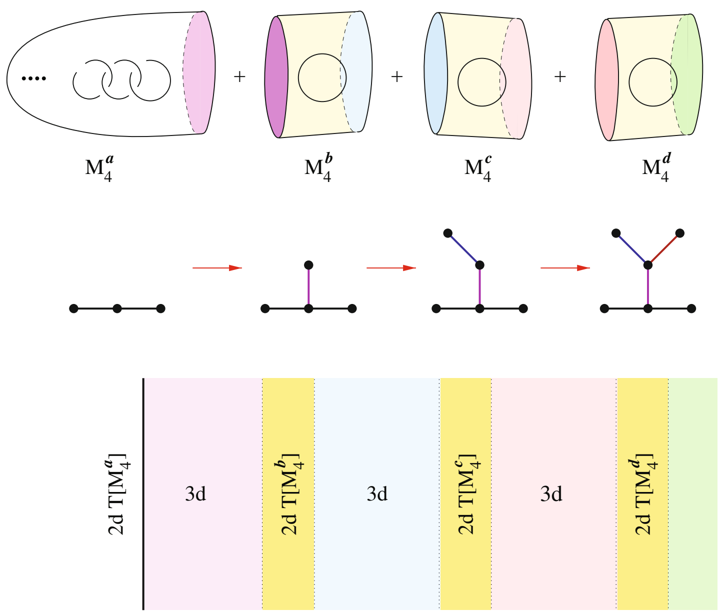


Fig. 15 The process of building a 4-manifold M_4 labeled by a plumbing tree can be represented by a sequence of basic cobordisms with $b_2 = 1$, where each step adds a new 2-handle. Each cobordism corresponds to a 2d wall (interface), and the process of building M_4 corresponds to defining $T[M_4]$ as the IR limit of the layered system of 3d theories trapped between walls shown on the lower part of the figure. Note, in general, there are many equivalent ways of building the same 4-manifold M_4 by attaching 2-handles in a different order; they correspond to equivalent descriptions (dualities) of the same 2d $(0,2)$ theory $T[M_4]$

$$\begin{array}{ccccccccc} 0 \to & H_2(B) & \to & H_2(M_4^+) & \to & H_2(M_4^+, B) & & \\ \| & \downarrow & & \downarrow & & \| & & \\ 0 \to & H_2(B, M_3^+) & \to & H_2(M_4^+, M_3^+) & \to & H_2(M_4^+, B) & \to & H_1(B, M_3^+) = 0 \\ & & & \wr\| \; \texttt{P.D.} & & \wr\| & & \\ & & & H^2(M_4^+) & & H_2(M_4^-, M_3^-) & & \\ & & & \wr\| & & \wr\| \; \texttt{P.D.} & & \\ & & & H_2(M_4^+)^* & \xrightarrow{\iota^*} & H_2(M_4^-)^* & & \end{array} \tag{93}$$

In this diagram, the map from $H_2(M_4^+)$ to its dual $H_2(M_4^+)^* \cong H^2(M_4^+)$ is given by the intersection form $Q^+ \equiv Q_{M_4^+}$. Therefore, we get

$$0 \to H_2(B) \to H_2(M_4^+) \xrightarrow{Q^+} H_2(M_4^+)^* \xrightarrow{\iota^*} H_2(M_4^-)^* \tag{94}$$

Since the second map, from $H_2(B)$ to $H_2(M_4^+)$, is injective, it follows that

$$H_2(B) \ = \ \ker\left(\iota^* \circ Q^+\right) \tag{95}$$

This useful result can tell us everything we want to know about the cobordism B from the data of M_4^- and M_4^+.

In particular, when both M_4^+ and M_4^- are sphere plumbings, and the plumbing tree of the former is obtained by adding a new vertex (with an edge) to the plumbing tree of the latter, as in Fig. 15, the second homology of the cobordism B is one-dimensional,

$$b_2(X) \ = \ 1\,, \tag{96}$$

and, therefore, its intersection form is determined by the self-intersection of a single generator $s \in H^2(B)$. Thus, introducing a natural basis $\{s_i\}$ for $H_2(M_4^+)$, such that the intersection pairing

$$Q^+(s_i, s_i) = Q^+_{ij} \tag{97}$$

is determined by the (weighted) plumbing tree, the generator $s \in H^2(B)$ can be expressed as a linear combination

$$s = \sum_{i=1}^{b_2(M_4^+)} k_i s_i \tag{98}$$

where the coefficients $k_i \in \mathbb{Z}$ are determined by (95):

$$Q^+(s,x) = 0\,, \qquad \forall x \in H_2(M_4^-) \tag{99}$$

In practice, of course, it suffices to verify this orthogonality condition only on the basis elements of $H_2(M_4^-)$. Then, it determines the cohomology generator (98) and, therefore, the self-intersection number $Q^+(s,s)$.

As a warm-up, let us illustrate how this works in the case of a linear plumbing in Fig. 1, where for simplicity we start with the case where all Euler numbers $a_i = -2$. Namely, if M_4^- has a linear plumbing graph with $n-1$ vertices and M_4^+ has a linear plumbing graph with n vertices, then the condition (99) becomes

$$Q(s, s_i) = 0\,, \qquad i = 1, \ldots, n-1 \tag{100}$$

or, more explicitly,

$$\begin{aligned} -2k_1 + k_2 &= 0 \\ k_{i-1} - 2k_i + k_{i+1} &= 0 \qquad i = 2, \ldots, n-1 \end{aligned} \tag{101}$$

Solving these equations we find the generator $s \in H^2(B)$,

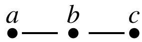
$$s = s_1 + 2s_2 + 3s_3 + \cdots + ns_n \tag{102}$$

for the cobordism B that relates A_{n-1} and A_n linear plumbings. Now, the self-intersection is easy to compute:

$$Q^+(s,s) = -n(n+1) \tag{103}$$

It is easy to generalize this calculation to linear plumbings with arbitrary framing coefficients a_i, as well as plumbing graphs which are not necessarily linear. As the simplest example of the latter, let us consider a 2-handle attachment in the first step of Fig. 15 that turns a linear plumbing graph with three vertices

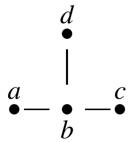
M_4^- :



(104)

into a non-linear plumbing graph with a trivalent vertex:

M_4^+ :



(105)

In order to determine the cobordism B that does the job we are again going to use (95) or, better yet, its more explicit version (99) suitable for arbitrary plumbing trees. As before, denoting by s_i the generators of $H_2(M_4^+)$ with the intersection pairing (97), which is easy to read off from (105), we get the system of linear equations (99) that determines the generator (98) of the cobordism B:

$$\begin{aligned} Q^+(s,s_1) &= ak_1 + k_2 \ = \ 0 \\ Q^+(s,s_2) &= k_1 + bk_2 + k_3 + k_4 \ = \ 0 \\ Q^+(s,s_3) &= k_2 + ck_3 \ = \ 0 \end{aligned} \tag{106}$$

Of course, in case of negative-definite 4-manifolds a, b, c, and d are all supposed to be negative. Solving these equations we find the integer coefficients in (98),

$$k_1 = \frac{c}{\gcd(a,c)}, \quad k_2 = -\frac{ac}{\gcd(a,c)}, \quad k_3 = \frac{a}{\gcd(a,c)}, \quad k_4 = \frac{abc-a-c}{\gcd(a,c)} \tag{107}$$

which, in turn, determine the intersection form on B:

$$Q^+(s,s) = \frac{(abcd - ac - ad - cd)(abc - a - c)}{\gcd(a,c)^2} \tag{108}$$

For instance, if $a = b = c = d = -2$, we get $Q_B = \langle -4 \rangle$.

3 Top-Down Approach: Fivebranes and Instantons

In this section we approach the correspondence between 4-manifolds and 2d $\mathcal{N} = (0, 2)$ theories $T[M_4; G]$ by studying the (flavored) elliptic genus (9) which, according to (10), should match the Vafa–Witten partition function.

In particular, we propose the "gluing rules" that follow operations on 4-manifolds introduced in Sect. 2 and identify the set of basic cobordisms with branching functions in certain coset models. In the non-abelian case, the key ingredient in the gluing construction is the integration measure, which we propose to be the index of a 2d $(0, 2)$ vector multiplet. Another key ingredient, which plays an important role in (10) for non-compact 4-manifolds, is a relation between *discrete basis* and *continuous basis* introduced in Sect. 3.10.

3.1 Vafa–Witten Theory

In order to realize the Vafa–Witten twist of 4d $\mathcal{N} = 4$ super-Yang-Mills [VW94] in M-theory, we start with the six-dimensional $(2, 0)$ theory realized on the world-volume of N fivebranes. The R-symmetry group of the $(2, 0)$ theory is $Sp(2)_r \cong SO(5)_r$ and can be viewed as a group of rotations in the five-dimensional space transverse to the fivebranes. A $(2, 0)$ tensor multiplet in six dimensions contains 5 scalars, 2 Weyl fermions and a chiral 2-form, which under $Sp(2)_r$ transform as **5**, **4**, and **1**, respectively.

We are interested in the situation when the M-theory space is $S^1 \times \mathbb{R}_t \times M_7 \times \mathbb{C}$, where M_7 is a 7-manifold with G_2 holonomy and $\mathbb{R}_t$ may be considered as the time direction. We introduce a stack of N fivebranes supported on the subspace $S^1 \times \mathbb{R}_t \times M_4$, where M_4 is a coassociative cycle in M_7. This means that the normal bundle of M_4 inside M_7 is isomorphic to the self-dual part of $\Lambda^2 T^* M_4$:

$$T_{M_7/M_4} \;\cong\; \Lambda^2_+ T^* M_4 \,. \tag{109}$$

Moreover, the neighborhood of M_4 in M_7 is isomorphic (as a G_2-manifold) to the neighborhood of the zero section of $\Lambda^2_+ T^* M_4$.

Since both the 11-dimensional space-time and the fivebrane world-volume in this setup have S^1 as a factor, we can reduce on this circle to obtain N D4-branes supported on $\mathbb{R} \times M_4$ in type IIA string theory. The D4-brane world-volume theory is maximally supersymmetric ($\mathcal{N} = 2$) super-Yang-Mills in five dimensions with the following field content:

spectrum of $5d$ super-Yang-Mills

	$Spin(5)_E$	$Sp(2)_r$
1-form	**5**	**1**
scalars	**1**	**5**
fermions	**4**	**4**

The rotation symmetry in the tangent bundle of M_4 is $Spin(4)_E \cong SU(2)_L \times SU(2)_R$ subgroup of the $Spin(5)_E$ symmetry of the Euclidean five-dimensional theory. Five normal direction to the branes are decomposed into three directions normal to M_4 inside M_7 and two directions of $\mathbb{C}$-plane. This corresponds to the following decomposition of the R-symmetry group:

$$SO(5)_r \;\rightarrow\; SO(3)_A \times SO(2)_U \cong SU(2)_A \times U(1)_U. \tag{110}$$

The fields of the 5d super-Yang-Mills transform under the resulting $SU(2)_L \times SU(2)_R \times SU(2)_A \times U(1)_U$ symmetry group as

$$\begin{aligned} \text{bosons}: &\quad (\mathbf{5},\mathbf{1}) \oplus (\mathbf{1},\mathbf{5}) \rightarrow (\mathbf{2},\mathbf{2},\mathbf{1})^0 \oplus (\mathbf{1},\mathbf{1},\mathbf{1})^0 \oplus (\mathbf{1},\mathbf{1},\mathbf{3})^0 \oplus (\mathbf{1},\mathbf{1},\mathbf{1})^{\pm 2} \\ \text{fermions}: &\quad (\mathbf{4},\mathbf{4}) \rightarrow (\mathbf{2},\mathbf{1},\mathbf{2})^{\pm 1} \oplus (\mathbf{1},\mathbf{2},\mathbf{2})^{\pm 1} \end{aligned} \tag{111}$$

Non-trivial embedding of the D4-branes in space-time with the normal bundle (109) corresponds [BVS95] to identifying $SU(2)_L$ with $SU(2)_A$ and gives precisely the topological twist introduced by Vafa in Witten [VW94]. The spectrum of the resulting theory looks like:

$$\begin{aligned} \text{bosons}: &\quad (\mathbf{2},\mathbf{2})^0 \oplus (\mathbf{1},\mathbf{1})^0 \oplus (\mathbf{3},\mathbf{1})^0 \oplus (\mathbf{1},\mathbf{1})^{\pm 2} \\ \text{fermions}: &\quad (\mathbf{1},\mathbf{1})^{\pm 1} \oplus (\mathbf{3},\mathbf{1})^{\pm 1} \oplus (\mathbf{2},\mathbf{2})^{\pm 1} \end{aligned} \tag{112}$$

where we indicate transformation under the symmetry group $SU(2)'_L \times SU(2)_R \times U(1)_U$. Here, the subgroup $SU(2)'_L \times SU(2)_R$ is the new rotation symmetry along M_4, whereas $U(1)_U$ is the R-symmetry[11] of the effective $\mathcal{N} = 2$ supersymmetric quantum mechanics $T_{1\mathrm{d}}[M_4]$ on $\mathbb{R}_t$. The $U(1)_U$ quantum number is called the ghost number.

From (112) it is clear that the resulting supersymmetric quantum mechanics $T_{1\mathrm{d}}[M_4]$ has two supercharges, which are scalar from the viewpoint of the 4-manifold M_4 and which carry ghost number $U = +1$ and $U = -1$, respectively. When the quantum mechanics is lifted to the 2d theory $T[M_4]$ on $S^1 \times \mathbb{R}_t$ they become supercharges of $\mathcal{N} = (0, 2)$ SUSY. Among the bosons, two states $(\mathbf{1},\mathbf{1})^{\pm 2}$ with non-zero ghost number are scalars ϕ and $\overline{\phi}$ that are not affected by the twist, the state $(\mathbf{3},\mathbf{1})^0$ is the self-dual 2-form field B, and finally the state $(\mathbf{1},\mathbf{1})^0$ is the scalar field

[11] Note, in [VW94] the symmetry group $U(1)_U$ is enhanced to the global symmetry group $SU(2)_U$ due to larger R-symmetry of the starting point.

C, all transforming in the adjoint representation of the gauge group. The state $(\mathbf{2},\mathbf{2})^0$ is, of course, the gauge connection on M_4:

$$\begin{array}{ll} (\mathbf{2},\mathbf{2})^0 & \text{gauge connection } A \\ (\mathbf{3},\mathbf{1})^0 & \text{self-dual 2-form } B \\ (\mathbf{1},\mathbf{1})^{\pm 2} & \text{complex scalar } \phi \\ (\mathbf{1},\mathbf{1})^0 & \text{real scalar } C \end{array} \tag{113}$$

Now let us consider a situation where the time direction is also compactified to a circle: $\mathbb{R}_t \rightsquigarrow S^1_t$ in a way that allows the M-theory circle S^1 to fiber non-trivially over S^1_t, so that the twisted product $S^1 \rtimes S^1_t$ is a torus with the complex modulus τ. Then, the theory on the fivebranes can be described as a theory on D4-branes supported on M_4, i.e., the four-dimensional topologically twisted $\mathcal{N} = 4$ SYM with coupling constant τ [VW94].

The path integral of the Vafa–Witten theory localizes on the solutions to the following equations:

$$\begin{aligned} F_A^+ - \frac{1}{2}[B \times B] + [C,B] = 0 \\ d_A^* B - d_A C = 0 \end{aligned} \qquad \text{where} \qquad \begin{aligned} &A \in \mathcal{G}_P \\ &B \in \Omega^{2,+}(M_4; \mathrm{ad}_P) \\ &C \in \Omega^0(M_4; \mathrm{ad}_P) \end{aligned} \tag{114}$$

where $\mathcal{G}_P$ denotes the space of connections of a principal bundle P. Under certain conditions (see [VW94] for details) the only non-trivial solutions are given by configurations with vanishing self-dual part of the curvature

$$F_A^+ \;=\; 0 \tag{115}$$

and trivial other fields ($B = 0$ and $d_A C = 0$). The partition function is then given by the generating function of the Euler numbers of instanton moduli spaces:

$$Z_{\mathrm{VW}}[M_4](q) \;=\; \sum_m \chi(\mathcal{M}_m) q^{m-\frac{c}{24}} \tag{116}$$

where

$$\mathcal{M}_m \;=\; \left\{ A \in \mathcal{G}_P \;:\; F_A^+ = 0,\; \langle \mathrm{ch}, [M_4] \rangle \equiv \frac{1}{8\pi^2} \int_{M_4} \mathrm{Tr}\, F^2 = m \right\} \;/\, \text{Gauge}\,,$$

$$q = e^{2\pi i \tau}$$

and c is a constant that depends on the topology of M_4. In [VW94] it was proposed that

$$c \;=\; N \cdot \chi(M_4) \tag{117}$$

where N is the rank of the gauge group and $\chi(M_4)$ is the Euler characteristic[12] of M_4. The constant c can be interpreted as the left-moving central charge c_L of the dual 2d $(0,2)$ theory $T[M_4]$.

In general, when the manifold M_4 is not compact and the gauge group is $U(N)$, anti-self-dual configurations can also be distinguished by the first Chern class c_1 and the boundary conditions at infinity. In order to have finite action, the connection should be asymptotically flat:

$$A|_{\partial M_4} = A_\rho, \qquad F_{A_\rho} = 0\,. \tag{118}$$

Therefore, as we already mentioned in Sect. 2.2, different asymptotics can be labeled by flat connections on the boundary 3-manifold $M_3 = \partial M_4$:

$$\rho \in \mathcal{M}_{\text{flat}}[M_3] \equiv \text{Hom}(\pi_1(M_3), U(N)) \,/\, \text{Gauge}\,. \tag{119}$$

The dependence on the first Chern class can be captured by introducing the following topological term in the action, cf. [DHSV07]:

$$\Delta S = \frac{1}{2\pi}\int_\xi \text{Tr}\, F \equiv \langle c_1, \xi\rangle \tag{120}$$

where $\xi \in H_2(M_4) \otimes \mathbb{C}$. It is useful to define the following exponential map:

$$\begin{aligned} \exp : H_2(M_4) \otimes \mathbb{C} &\longrightarrow (\mathbb{C}^*)^{b_2} \\ \xi &\longmapsto x \end{aligned} \tag{121}$$

such that $\ker(\exp) = H_2(M_4, \mathbb{Z})$ and also the "power" operation

$$\begin{aligned} (\mathbb{C}^*)^{b_2} \times H^2(M_4) &\longrightarrow \mathbb{C}^* \\ (x, h) &\longmapsto x^h \equiv e^{2\pi i \langle h, \xi\rangle} \end{aligned} \tag{122}$$

for some preimage ξ of x. The refined Vafa–Witten partition function then depends on $b_2(M_4)$ additional fugacities and is given by

$$Z_{\text{VW}}[M_4]_\rho(q, x) = \sum_{m, c_1} \chi(\mathcal{M}_{m, c_1, \rho})\, q^{m - \frac{c}{24}}\, x^{c_1} \tag{123}$$

where

$$\begin{aligned} \mathcal{M}_{m,c_1,\rho} = \big\{ &A \in \mathcal{G}_P \;:\; F_A^+ = 0,\; \langle \text{ch}, [M_4]\rangle \\ &= m,\; [\text{Tr}\, F] = 2\pi c_1,\; A|_{M_3} = A_\rho \big\} \,/\text{Gauge}. \end{aligned}$$

[12]When M_4 is non-compact $\chi(M_4)$ should be replaced by the regularized Euler characteristic, and when $G = U(N)$ one needs to remove by hand the zero-mode to ensure that the partition function does not vanish identically.

From the point of view of the 2d theory $T[M_4; U(N)]$, the fugacities x in (123) play the role of flavor fugacities in the elliptic genus. This tells us that $T[M_4; U(N)]$ has flavor symmetry of rank b_2 associated with 2-cycles of M_4.

In what follows, if not explicitly stated otherwise, we will consider 4-manifolds (13) with

$$\begin{gathered} b_2^+(M_4) = 0\,, \qquad \pi_1(M_4) = 0\,, \qquad H_2(M_3, \mathbb{Z}) = 0\,, \qquad H_1(M_3, \mathbb{R}) = 0 \\ \Gamma \equiv H_2(M_4, \mathbb{Z}) \cong \mathbb{Z}^{b_2}\,, \qquad \Gamma^* \equiv H^2(M_4, \mathbb{Z}) \cong \mathbb{Z}^{b_2} \end{gathered} \tag{124}$$

The last two conditions mean that there is no torsion in second (co)homology. As explained in Sect. 2.1, such manifolds are uniquely defined by the intersection form or, alternatively, by the plumbing graph.

3.2 Gluing Along 3-Manifolds

In this section we will describe how the Vafa–Witten partition function behaves under cutting and gluing of 4-manifolds. Suppose one can produce a 4-manifold M_4 by gluing M_4^+ and M_4^- along a common boundary component M_3. For simplicity, in the following we actually assume that M_3 is the only boundary component for both M_4^+ and M_4^- (that is, the resulting manifold M_4 does not have any boundary). The generalization to the case when $M_4^\pm$ have other boundary components (that will become boundary components of M_4 after the gluing) is straightforward. For the same reason we will also suppress the dependence of the moduli spaces on the first Chern class c_1 or, equivalently, the dependence of the Vafa–Witten partition function on the fugacities x in (123).

Since for $b_2^+ > 1$ we expect the topology of the instanton moduli spaces to be independent under smooth deformations of the 4-manifold, consider the situation where the boundary neighborhoods of $M_4^\pm$ look like long "half-necks" of the form $\mathbb{R}_+ \times M_3$, as illustrated in Fig. 16. Very naively the Vafa–Witten partition function on M_4 is given by a sum of products of partition functions on $M_4^\pm$ with identified

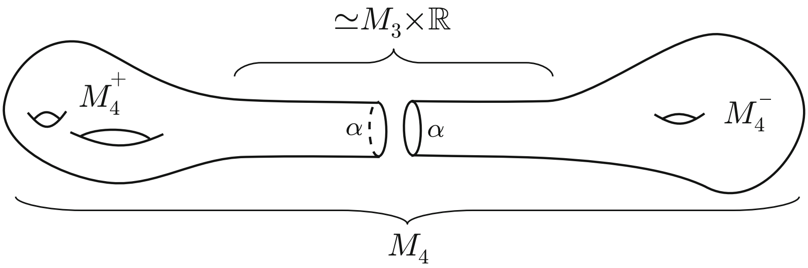


Fig. 16 Gluing of M_4^+ and M_4^- along the common boundary M_3

boundary conditions. However in this way we count instantons living on the long neck $M_3 \times \mathbb{R}$ twice and we need to cancel out this contribution.

Let us address this issue more systematically. Let $\widetilde{\mathcal{M}}_m^{\alpha\beta}$ be the moduli space of m instantons[13] on $M_3 \times \mathbb{R}$ with boundary conditions $\alpha, \beta \in \mathcal{M}_{\text{flat}}[M_3]$. One can always factor out the part of the moduli space associated with translations along $\mathbb{R}$:

$$\widetilde{\mathcal{M}}_m^{\alpha\beta} = \overline{\mathcal{M}}_m^{\alpha\beta} \times \mathbb{R}. \tag{125}$$

Let us denote the corresponding generating function for Euler characteristics as follows:

$$K^{\alpha\beta}[M_3] \equiv \sum_m \chi(\widetilde{\mathcal{M}}_m^{\alpha\beta}) q^m. \tag{126}$$

Now let $\mathcal{M}_m$ and $\mathcal{M}_{m,\alpha}^{\pm}$ be instanton moduli spaces for M_4 and $M_4^{\pm}$, respectively. Then

$$\mathcal{M}_m = \bigcup_{\substack{\alpha \\ m_+ + m_- = m}} \mathcal{M}_{m_+,\alpha}^{+} \times \mathcal{M}_{m_-,\alpha}^{-}. \tag{127}$$

The problem, however, is that this union is *not* disjoint. Various terms have common boundary components corresponding to particular degeneration of instanton configurations. Common codimension-1 boundary components have the following form:

$$\mathcal{M}_{m_+,\alpha}^{+} \times \overline{\mathcal{M}}_{\Delta}^{\alpha\beta} \times \mathcal{M}_{m_-,\beta}^{-} \subset \begin{array}{c} \partial\left(\mathcal{M}_{m_+ + \Delta,\beta}^{+} \times \mathcal{M}_{m_-,\beta}^{-}\right) \\ \text{and} \\ \partial\left(\mathcal{M}_{m_+,\alpha}^{+} \times \mathcal{M}_{\Delta + m_-,\alpha}^{-}\right). \end{array} \tag{128}$$

The first case can be intuitively understood from a limit when we separate a localized configuration with instanton number Δ in M_4^+ and push it to the boundary. And in the second case we do the same for M_4^-. Similarly, there are common codimension-2 boundary components:

$$\mathcal{M}_{m_+,\alpha}^{+} \times \overline{\mathcal{M}}_{\Delta_1}^{\alpha\beta} \times \overline{\mathcal{M}}_{\Delta_2}^{\beta\gamma} \times \mathcal{M}_{m_-,\gamma}^{-} \subset \begin{array}{c} \partial\left(\mathcal{M}_{m_+ + \Delta_1 + \Delta_2,\gamma}^{+} \times \mathcal{M}_{m_-,\gamma}^{-}\right) \\ \partial\left(\mathcal{M}_{m_+ + \Delta_1,\beta}^{+} \times \mathcal{M}_{\Delta_2 + m_-,\beta}^{-}\right) \\ \partial\left(\mathcal{M}_{m_+,\alpha}^{+} \times \mathcal{M}_{\Delta_1 + \Delta_2 + m_-,\alpha}^{-}\right) \end{array} \tag{129}$$

and so on.

[13] Here and in what follows the instanton number is not necessarily an integer.

Then, applying inclusion–exclusion principle for Euler characteristic we get

$$\chi(\mathcal{M}_m) = \sum_{\substack{\alpha \\ m_+ + m_- = m}} \chi\left(\mathcal{M}^+_{m_+,\alpha} \times \mathcal{M}^-_{m_-,\alpha}\right) - \sum_{\substack{\alpha,\beta;\ \Delta>0 \\ m_+ + \Delta + m_- = m}} \chi\left(\mathcal{M}^+_{m_+,\alpha} \times \overline{\mathcal{M}}^{\alpha\beta}_{\Delta} \times \mathcal{M}^-_{m_-,\beta}\right) + \sum_{\substack{\alpha,\beta,\gamma;\ \Delta_{1,2}>0 \\ m_+ + \Delta_1 + \Delta_2 + m_- = m}} \chi\left(\mathcal{M}^+_{m_+,\alpha} \times \overline{\mathcal{M}}^{\alpha\beta}_{\Delta_1} \times \overline{\mathcal{M}}^{\beta\gamma}_{\Delta_2} \times \mathcal{M}^-_{m_-,\gamma}\right) - \cdots \tag{130}$$

which translates into the following relation for the generating functions:

$$\begin{aligned} Z_{\mathrm{VW}}[M_4] &= \sum_\alpha Z_{\mathrm{VW}}[M_4^+]_\alpha Z_{\mathrm{VW}}[M_4^-]_\alpha - \sum_{\alpha,\beta} Z_{\mathrm{VW}}[M_4^+]_\alpha (K^{\alpha\beta}[M_3] - \delta^{\alpha\beta}) Z_{\mathrm{VW}}[M_4^-]_\beta \\ &\quad + \sum_{\alpha,\beta,\gamma} Z_{\mathrm{VW}}[M_4^+]_\alpha (K^{\alpha\beta}[M_3] - \delta^{\alpha\beta})(K^{\beta\gamma}[M_3] - \delta^{\beta\gamma}) Z_{\mathrm{VW}}[M_4^-]_\gamma - \cdots \\ &= \sum_{\alpha,\beta} Z_{\mathrm{VW}}[M_4^+]_\alpha (K^{-1}[M_3])^{\alpha\beta} Z_{\mathrm{VW}}[M_4^-]_\beta \end{aligned} \tag{131}$$

where $K^{-1}[M_3]$ denotes the matrix inverse to $K[M_3]$ defined in (126). The relation (131) obviously holds when $M_4 = M_4^+ = M_4^- = M_3 \times \mathbb{R}$. Let us note that in the case when M_3 is a lens space the "gluing kernel" $K[M_3]$ can be explicitly computed using the results of [Aus90, FH90].

For later convenience, let us define a modified Vafa–Witten partition with an *upper* index denoting the boundary condition:

$$Z_{\mathrm{VW}}[M_4^-]^\alpha \equiv \sum_\beta (K^{-1}[M_3])^{\alpha\beta} Z_{\mathrm{VW}}[M_4^-]_\beta. \tag{132}$$

Intuitively this modification can be understood as excluding instantons approaching the boundary. Then the relation between partition functions takes the following simple form:

$$Z_{\mathrm{VW}}[M_4] = \sum_\alpha Z_{\mathrm{VW}}[M_4^+]_\alpha Z_{\mathrm{VW}}[M_4^-]^\alpha. \tag{133}$$

3.3 Relation to Affine Lie Algebras

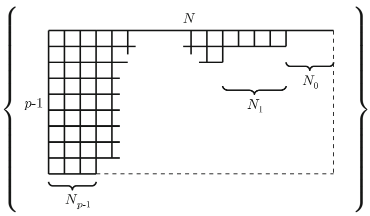
Before we discuss cobordisms, let us review the relation between Vafa–Witten theory on ALE spaces and affine Lie algebras [Nak94, VW94, DHSV07], that will be our starting point for constructing generalizations. Namely, let M_4 be a hyper-Kähler ALE space obtained by a resolution of the quotient singularity $\mathbb{C}^2/H$, where H is a finite subgroup of $SU(2)$. According to the McKay correspondence, finite subgroups of $SU(2)$ have ADE classification and therefore for each H there is a corresponding simple Lie algebra $\mathfrak{g}$ of the same ADE type. From the work of Nakajima [Nak94] it follows that the partition function of the Vafa–Witten theory with the gauge group $U(N)$ is given by the character of the integrable representation of the corresponding affine Lie algebra $\hat{\mathfrak{g}}$ at level N:

$$Z_{\text{VW}}^{U(N)}[M_4]_\rho(q,x) \;=\; \chi_\rho^{\hat{\mathfrak{g}}_N}(q,x)\,. \tag{134}$$

Let us explain in some detail the role of the parameters ρ, q, and x on the right-hand side of this formula. First, the formula (134) exploits the fact that there is a one-to-one correspondence between $U(N)$ flat connections on $M_3 \cong S^3/H$ and integrable representations of $\hat{\mathfrak{g}}_N$. The right-hand side of (134) can then be understood as a character of $\hat{\mathfrak{g}}_N$ for a given representation ρ. Let us consider how the identification between flat connections and integrable representations works in a simple case when $H = \mathbb{Z}_p$, $M_4 = A_{p-1}$ and $\mathfrak{g} = \mathfrak{su}(p)$. The set of flat connections (119) in this case is given by the ordered partitions of N with p parts, which are in one-to-one correspondence with Young diagrams that have at most $p-1$ rows and N columns:

$$\mathrm{Hom}(\mathbb{Z}_p, U(N))/U(N) = \left\{ \begin{pmatrix} z_1 & & 0 \\ & \ddots & \\ 0 & & z_N \end{pmatrix}^p = 1 \right\} / S_N =$$

$$\left\{ \mathrm{diag}(\underbrace{1,\ldots,1}_{N_0}, \underbrace{e^{\frac{2\pi i}{p}},\ldots,e^{\frac{2\pi i}{p}}}_{N_1},\ldots,\underbrace{e^{2\pi i\frac{p-1}{p}},\ldots,e^{2\pi i\frac{p-1}{p}}}_{N_{p-1}}) \right\} \cong$$



(135)

Young diagrams of such type indeed describe integrable representation of $\widehat{\mathfrak{su}}(p)_N$. The variables (q,x) in the right-hand side of (134) play the role of coordinates on the (complexified) torus corresponding to the Cartan subalgebra $\hat{\mathfrak{h}}$ of $\hat{\mathfrak{g}}_N$. In particular, τ is a coordinate on $\hat{\mathfrak{h}}$ in the direction of L_0 and x can be interpreted as coordinates on the (complexified) maximal torus of the Lie group G corresponding to the ordinary Lie algebra $\mathfrak{g}$. This is in agreement with the fact that the lattice Γ^* for an ALE space of the *ADE* type is the same as the weight lattice of the corresponding simple Lie algebra $\mathfrak{g}$ and ξ in (122) is then the element of the dual space. The dual lattice Γ is the same as the root lattice of $\mathfrak{g}$ and the intersection form Q plays the role of the normalized Killing form. It follows that the abelian quiver CS with coefficients Q_{ij} is the same as the ordinary CS with the gauge group G restricted to the Cartan subalgebra, which can be interpreted as a level-rank duality.

Now let us describe the gluing of 4-manifolds considered in Sect. 2.7 in the language of (affine) Lie algebras. Suppose the manifold M_4^+ with boundary M_3^+ is defined by a plumbing graph of *ADE* type which can be interpreted as a Dynkin diagram of Lie algebra $\mathfrak{g}^+$ with root lattice $\Gamma_+ \equiv H_2(M_4^+)$. Let us pick up a subalgebra $\mathfrak{g}^- \subset \mathfrak{g}^+$ and consider the manifold M_4^+ with properties (124) such that the lattice $\Gamma_- \equiv H_2(M_4^-)$ is the root lattice of $\mathfrak{g}^-$. The lattice Γ_- is a sublattice of Γ_+ and the manifold M_4^+ can be obtained by gluing M_4^- with a certain (co)bordism B such that $\partial B = M_3^- \sqcup M_3^+$ along the common boundary component M_3^-, cf. (92). In the rest of the paper we will sometimes use the following schematic (but intuitive) notation for the process of obtaining a manifold M_4^+ by gluing a cobordism B to M_4^-:

$$M_4^- \overset{B}{\rightsquigarrow} M_4^+ . \tag{136}$$

From the gluing principle described in the previous section we have:

$$Z_{\text{VW}}^{U(N)}[M_3^+]_\rho(q,x) \;=\; \sum_\lambda Z_{\text{VW}}^{U(N)}[B]_\rho^\lambda(q,x^\perp)\, Z_{\text{VW}}^{U(N)}[M_3^-]_\lambda(q,x^\parallel) \tag{137}$$

where the splitting of the parameters $x = (x^\perp, x^\parallel)$ corresponds to the splitting[14] of the homology groups $H_2(M_4^+) \otimes \mathbb{C} = H_2(B) \otimes \mathbb{C} \oplus H_2(M_4^-) \otimes \mathbb{C}$. Using (134) one has

$$\chi_\rho^{\hat{\mathfrak{g}}_N^+}(q,x) \;=\; \sum_\lambda Z_{\text{VW}}^{U(N)}[B]_\rho^\lambda(\tau,x^\perp)\; \chi_\lambda^{\hat{\mathfrak{g}}_N^-}(q,x^\parallel)\,. \tag{138}$$

Therefore, $Z_{\text{VW}}^{U(N)}[B]_\rho^\lambda$ are given by the branching functions of the embedding $\mathfrak{g}^- \subset \mathfrak{g}^+$,

$$\boxed{Z_{\text{VW}}^{U(N)}[B]_\rho^\lambda \;=\; \chi_{\lambda,\rho}^{\hat{\mathfrak{g}}_N^+/\hat{\mathfrak{g}}_N^-}} \tag{139}$$

[14] Let us note that $H_2(M_4^+) \neq H_2(B) \oplus H_2(M_4^-)$. However, the lattice $H_2(M_4^+)$ can be obtained from the lattice $H_2(B) \oplus H_2(M_4^-)$ by the so-called gluing procedure that will be described in detail shortly.

Table 2 Dictionary between Vafa–Witten theory and (affine) Lie algebras

Physics and geometry	Algebra
Plumbing graph	Dynkin diagram of $\mathfrak{g}$
Fugacities x	Maximal torus of G
Coupling τ	Coordinate on $\hat{\mathfrak{h}}$ along L_0
Intersection form	Normalized Killing form of $\mathfrak{g}$
$b_2(M_4)$	Rank of $\mathfrak{g}$
$H_2(M_4)$	Root lattice of $\mathfrak{g}$
$H^2(M_4)$	Weight lattice of $\mathfrak{g}$
Boundary condition	Integrable representation of $\hat{\mathfrak{g}}$
Rank of the gauge group	Level of $\hat{\mathfrak{g}}$
$Z_{\text{VW}}[M_4]$	Character of $\hat{\mathfrak{g}}$
Cobordism B: $M_4^+ = B \cup M_4^-$	Embedding $\mathfrak{g}^- \subset \mathfrak{g}^+$
$Z_{\text{VW}}[B]$	Branching functions

Let us consider a particular example: $M_4^+ = A_p$ and $M_4^- = A_{p-1}$. As was shown in Sect. 2.14 via a variant of the "Norman trick" [Nor69, Qui79], the cobordism B in this case is a 4-manifold in family (124) with a single 2-cycle of self-intersection $-(p+1)p$ and the boundary $L(p,-1) \sqcup L(p+1,-1)$. The partition function on B is then given by the characters of $\mathfrak{su}(p+1)/\mathfrak{su}(p)$ cosets:

$$Z_{\text{VW}}^{U(N)}[B]_\rho^\lambda \;=\; \chi_{\lambda,\rho}^{\widehat{\mathfrak{su}}(p+1)_N/\widehat{\mathfrak{su}}(p)_N}\,. \tag{140}$$

The relation between Vafa–Witten theory and (affine) Lie algebras is summarized in Table 2 and will play an important role in the following sections. In the next section we consider in detail the case of the gauge group $U(1)$. Then, in Sect. 3.9, we will make some proposals about the non-abelian case.

3.4 Cobordisms and Gluing in the Abelian Case

For a 4-manifold M_4 that satisfies (124) one has the short exact sequence (29):

$$0 \longrightarrow H_2(M_4) \xrightarrow{Q} H^2(M_4) \xrightarrow{i_{M_3}^*} H^2(M_3) \longrightarrow 0 \tag{141}$$

where the map Q is given by the intersection matrix and i_{M_3} is the inclusion map of the boundary $M_3 = \partial M_4$ into M_4. Equivalently, $H^2(M_3) \cong \text{coker} Q$.

In the case of abelian theory self-duality condition implies that

$$dF = 0\,, \qquad d^*F = 0. \tag{142}$$

For manifolds with asymptotically cylindrical or conical ends it has been shown (under certain assumptions) [APS73, Loc87] that the space of L^2 integrable 2-forms satisfying conditions (142) coincides with the space harmonic 2-forms $\mathcal{H}^2(M_4)$ and is isomorphic to the image of the natural map $H^2(M_4, M_3, \mathbb{R}) \longrightarrow H^2(M_4, \mathbb{R})$. In our case this map is an isomorphism. Since $b_2^+(M_4) = 0$ the space $\mathcal{H}^2(M_4)$ is an eigenspace of the Hodge $*$ operator with eigenvalue -1.

For an ordinary $U(1)$ gauge theory the Dirac quantization condition implies that $[F/2\pi] \in H^2(M_4) \equiv \Gamma^*$. However, since we are interested in gauge theory on the world-volume of a D4-brane in type IIA string theory setup, we need to take into account the Freed–Witten anomaly [FW99]. Specifically, the two-form $F = dA$ should be viewed as a curvature of the $U(1)$ part of a connection on a $Spin^c(4) \equiv Spin(4) \times_{\mathbb{Z}_2} U(1)$ principal bundle over M_4 obtained by a lift of the $SO(4)$ orthonormal frame bundle. Let us note that such a lift is possible for any 4-manifold, i.e., any 4-manifold is Spin^c. Not any 4-manifold, though, has a Spin structure. The obstruction is given by the second Stiefel–Whitney class $w_2 \in H_2(M_4, \mathbb{Z}_2)$. Therefore, as in [GVW00, GST02] we have a shifted quantization condition for the magnetic flux through a 2-cycle $C \subset M_4$:

$$\int_C \frac{F}{2\pi} = \frac{1}{2} \int_C w_2 = \frac{1}{2} Q(C, C) \mod \mathbb{Z} \tag{143}$$

where the second equality is the Wu's formula. The class $[F/2\pi]$ then takes values in the shifted lattice:

$$\left[\frac{F}{2\pi} \right] \in \widetilde{\Gamma}^* \equiv \Gamma^* + \Delta \tag{144}$$

where 2Δ is a lift[15] of w_2 with respect to the map $\Gamma^* \equiv H^2(M_4, \mathbb{Z}) \to H^2(M_4, \mathbb{Z}_2)$. From the Wu's formula it follows that $w_2 = 0$ or, equivalently, the manifold M_4 is Spin, if and only if the lattice Γ is even.

Let us note that since $\pi_1(M_4) = 0$ there are no non-trivial flat connections and therefore fixing $[F/2\pi]$ in $\widetilde{\Gamma}^*$ completely determines the anti-self-dual gauge connection. On the boundary $F|_{M_3} = 0$ and therefore $A|_{M_3}$ is a flat connection on M_3 which determines $[F/2\pi]$ modulo $H^2(M_4, M_3) \equiv \Gamma$. It is easy to see that the coset space $\widetilde{\Gamma}^*/\Gamma$ coincides with the space of flat connections. From (141) it follows that $H_1(M_3)$ is a finite abelian group of order $|\det Q|$. All such groups are isomorphic to a direct sum of finite cyclic groups. Therefore the space of flat connections on the boundary is given by

$$\text{Hom}(\pi_1(M_3), U(1)) \cong \text{Hom}(H_1(M_3), U(1)) \cong H^2(M_3) \cong \Gamma^*/\Gamma \cong \widetilde{\Gamma}^*/\Gamma \tag{145}$$

where the last equality follows from (141) and (144).

[15] Such lift exists because the manifold is Spin^c.

The Vafa–Witten partition for $U(1)$ gauge group can be calculated explicitly for general 4-manifold M_4 in the family (124) for a prescribed boundary condition $\rho \in \widetilde{\Gamma^*/\Gamma}$ and a fugacity $x \in H_2(M_4, \mathbb{R})$, cf. [Wit96, DVV02]:

$$\begin{aligned} Z_{\text{VW}}^{U(1)}[M_4]_\rho(q,x) &= \frac{1}{\eta^{\chi(M_4)}(q)} \sum_{\substack{[F/2\pi]\in\widetilde{\Gamma^*} \\ [F/2\pi]=\rho \mod \Gamma}} q^{\frac{1}{8\pi^2}\int F\wedge F} x^{[F/2\pi]} \\ &= \frac{1}{\eta^{\chi(M_4)}(q)} \sum_{\substack{[F/2\pi]\in\widetilde{\Gamma^*} \\ [F/2\pi]=\rho \mod \Gamma}} q^{-\frac{1}{2}Q^{-1}([F/2\pi],[F/2\pi])} x^{[F/2\pi]} \\ &= \frac{1}{\eta^{\chi(M_4)}(q)} \sum_{\gamma\in\Gamma} q^{-\frac{1}{2}Q^{-1}(Q\gamma+\rho,Q\gamma+\rho)} x^{Q\gamma+\rho} \\ &= \frac{1}{\eta^{\chi(M_4)}(q)} \sum_{\gamma\in\Gamma} q^{-\frac{1}{2}Q(\gamma+Q^{-1}\rho,\gamma+Q^{-1}\rho)} x^{Q\gamma+\rho}. \end{aligned} \tag{146}$$

The overall factor

$$\frac{1}{\eta^{\chi(M_4)}(q)} = q^{-\frac{\chi(M_4)}{24}} \sum_{m=0}^{\infty} \chi(\text{Hilb}^{[m]}(M_4))\, q^m \tag{147}$$

is the contribution of point-like instantons. Let us remind that the moduli space of m point-like instantons is given by the Hilbert scheme $\text{Hilb}^{[m]}(M_4)$ which can be understood as a regularization of the configuration space of m points on M_4.

Since the quadratic form $-Q$ is positive definite one can always assume that the lattices Γ and Γ^* are embedded in the Euclidean space $\mathbb{R}^{b_2}$ so that

$$\Gamma^* = \{n_i\omega_i | n_i \in \mathbb{Z}\} \subset \mathbb{R}^{b_2}\,.$$

and

$$\Gamma = \{n_i\lambda_i | n_i \in \mathbb{Z}\} \subset \Gamma^* \subset \mathbb{R}^{b_2}$$

The basis vectors of these lattices are chosen so that $(\lambda_i, \lambda_j) = -Q_{ij}$ and $(\omega_i, \lambda_j) = \delta_{ij}$ where $(\cdot,\cdot)$ is the standard Euclidean scalar product. The shift due to the Freed–Witten anomaly can be represented then by the vector $\Delta = \frac{1}{2}\sum_i \|\lambda_i\|^2\omega_i$. In this setup (146) reads simply as:

$$\begin{aligned} Z_{\text{VW}}^{U(1)}[M_4]_\rho(q,x) &= \frac{1}{\eta^{\chi(M_4)}(q)} \sum_{\gamma\in\Gamma\subset\mathbb{R}^{b_2}} q^{\frac{1}{2}\|\gamma+\rho+\Delta\|^2} x^{\gamma+\rho+\Delta} \\ &\equiv \frac{\theta_\Gamma^{(\rho+\Delta)}(x;q)}{\eta^{\chi(M_4)}(q)}, \qquad \rho\in\Gamma^*/\Gamma. \end{aligned} \tag{148}$$

where $\theta_\Gamma^{(\rho+\Delta)}$ stands for the theta function of the lattice Γ with the shift $\rho + \Delta$. The regularized Euler characteristic $\chi(M_4)$ coincides with dimension of the lattice b_2.

3.5 Number of Vacua

As in [GVW00, GST02], the quantum mechanics $T_{1\text{d}}[M_4]$ on $\mathbb{R}_t$ obtained by reduction of an M5-brane on $S^1 \times M_4$ is specified by a flat connection A_ρ on the boundary and the flux at infinity which, up to constant depending on the topology of $M_4 \subset M_7$, is given by

$$\Phi_\infty = N_{D0} - \frac{1}{8\pi^2} \int_{M_4} F \wedge F \tag{149}$$

Here, N_{D0} is a non-negative integer denoting the number of point-like instantons. The origin of the last term is the Wess–Zumino part of the D4-brane action:

$$I_{WZ} = -\int_{\mathbb{R}\times M_4} C_* \wedge \text{ch}(F) \wedge \sqrt{\frac{\widehat{A}(TM_4)}{\widehat{A}(NM_4)}}. \tag{150}$$

Once we picked Φ_∞ and fixed the value of $[F/2\pi]$ modulo Γ which specify the theory $T_{1\text{d}}[M_4]_{\rho,\Phi_\infty}$, its supersymmetric vacua are obtained by finding $N_{D0} \geq 0$ and $[F/2\pi]$ which solve (149). Note, the effective theory is massive when $N_{D0} = 0$. If $N_{D0} > 0$ there are moduli of point-like abelian on M_4. The number of vacua is given by the corresponding coefficient of (123):

$$\#\{\text{vacua of } T_{1\text{d}}[M_4]_{\rho,\Phi_\infty}\} \;=\; Z_{\text{VW}}[M_4]_\rho(q,0)\big|_{\text{coefficient of } q^{\Phi_\infty - \frac{c}{24}}} \tag{151}$$

Let us consider $M_4 = A_{p-1}$ as an example. The lattice Γ is even in this case and therefore $\widetilde{\Gamma}^* = \Gamma^*$. As was mentioned earlier, Γ and Γ^* can be interpreted as the root and weight lattices of $\mathfrak{su}(p)$. These lattices can be naturally embedded into $\mathbb{R}^{p-1}$, which in turn can be considered as the subspace of $\mathbb{R}^p$ orthogonal to the vector $(1, \ldots, 1)$. The root lattice can be generated by simple roots satisfying $\|\lambda_i\|^2 = 2$ and $(\lambda_i, \lambda_{i+1}) = -1$. The weight lattice can be generated by $\omega_r,\ r = 1, \ldots, p-1$, the highest weights of the fundamental representations which can be realized as $\Lambda^r\mathbb{C}^p$. Let us also define $\omega_0 \equiv 0$. In the coset $\Gamma^*/\Gamma \cong \mathbb{Z}_p$ one has $\omega_r \sim r\omega_1$. For a given boundary condition $r = 0, \ldots, p-1$ the flux at infinity has the following form:

$$\Phi_\infty \;=\; N_{\text{D0}} + \frac{1}{2}\left\| \sum_{i=1}^{p-1} n_i\lambda_i + \omega_r \right\|^2, \qquad n_i \in \mathbb{Z}\,. \tag{152}$$

The massive vacua of the theory $T_{\text{1d}}[A_{p-1}]_{\rho,\Phi_\infty}$ correspond to the weights $w = \sum_{i=1}^{p-1} n_i\lambda_i + \omega_r$ that minimize (152) when $N_{\text{D0}} = 0$. The set of such weights is precisely the set of weights of the fundamental representation of $\mathfrak{su}(p)$ with the highest weight ω_r. Therefore one has

$$\#\{\text{vacua of } T_{\text{1d}}[A_{p-1}]_r\} \;=\; \dim \Lambda^r\mathbb{C}^p \;=\; \frac{p!}{r!(p-r)!}. \tag{153}$$

Up to a permutation, these weights have the following coordinates:

$$w \underset{S_p}{\sim} (\underbrace{1-\frac{r}{p},\ldots,1-\frac{r}{p}}_{r},\underbrace{-\frac{r}{p},\ldots,-\frac{r}{p}}_{p-r}). \tag{154}$$

The minimal value of the flux at infinity equals then

$$\Phi_\infty \;=\; \frac{(p-r)r}{2p}\,. \tag{155}$$

3.6 Gluing in the Abelian Case

Consider two 4-manifolds (not necessarily connected) $M_4^\pm$, both satisfying (124), with boundaries $\partial M_4^\pm = M_3^\pm$. Let us denote $\Gamma_\pm \equiv H_2(M_4^\pm)$ and $T_\pm \equiv H^2(M_3^\pm) \cong H_1(M_3^\pm)$ so that

$$0 \longrightarrow \Gamma_\pm \hookrightarrow \Gamma_\pm^* \xrightarrow{\pi_\pm} T_\pm \longrightarrow 0. \tag{156}$$

Suppose that M_4^+ can be obtained from M_4^- by gluing to the latter a certain (co)bordism B with boundary $\partial B = -M_3^- \sqcup M_3^+$.

Also, let us suppose that $b_2(B) = 0$ and the torsion groups in the long exact sequence (27) for the pair $(B,\partial B)$ are $T_2 = 0$ and $T_1 \equiv T$. This means that the only non-trivial cohomology of B and ∂B is contained in the following *finite* groups:

$$H_2(B,\partial B) \cong H^2(B) = T \tag{157}$$

$$H_1(B) \cong H^3(B,\partial B) = T \tag{158}$$

$$H_1(\partial B) \cong H^2(\partial B) = T_- \oplus T_+ \tag{159}$$

The sequence (27) then reduces to the following short exact sequence of finite abelian groups:

$$0 \longrightarrow T \xrightarrow{\upsilon=\upsilon_-\oplus\upsilon_+} T_-\oplus T_+ \xrightarrow{\psi} T \longrightarrow 0 \tag{160}$$

Let us denote the family of all such “basic” cobordisms by $\mathfrak{B}$. From the Mayer–Vietoris sequence for the pair of manifolds M_4^- and B glued along M_3^- one can deduce the following commutative diagram:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \Gamma_+^* & \longrightarrow & \Gamma_-^* \oplus T & \longrightarrow & T_- & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \pi_+} & & \downarrow{\scriptstyle (\pi_- - \upsilon_-)\oplus \upsilon_+} & & \downarrow{\scriptstyle \mathrm{id}} & & \\ 0 & \longrightarrow & T_+ & \longrightarrow & T_- \oplus T_+ & \longrightarrow & T_- & \longrightarrow & 0 \end{array} \tag{161}$$

where both horizontal lines form short exact sequences. From the snake lemma it follows that $\Gamma_+ = \ker \pi_+$ can be realized as a sublattice of Γ_-^*:

$$\begin{aligned} \Gamma_+ &= \ker(\pi_- - \upsilon_-) \oplus \upsilon_+ = \pi_-^{-1}\left[\mathrm{im}\, \upsilon_-|_{\ker \upsilon_+}\right] \\ &= \{\alpha \in \Gamma_-^* \mid \exists \rho \in T \ \text{ s.t. } \ \alpha \mod \Gamma_- = \upsilon_-(\rho),\ \upsilon_+(\rho) = 0\}\,. \end{aligned} \tag{162}$$

Let us now briefly review the notion of gluing of lattices described in detail in, e.g., [GL91]. Consider some integer lattice Γ embedded into a Euclidean space and a finite family of *glue vectors* $g_i \in \Gamma^*$. Then one can define the *glued lattice*

$$\Gamma' = \{\gamma + \sum_i n_i g_i \mid \gamma \in \Gamma,\ n_i \in \mathbb{Z}\} \subset \Gamma^*. \tag{163}$$

The finite abelian group $J \equiv \Gamma'/\Gamma$ is called the *glue group*. It is a subgroup of Γ^*/Γ generated by the equivalence classes $[g_i]$. As was considered in detail in [GL92, GL92], the gluing operation produces identities on the corresponding theta functions defined as in (148):

$$\theta_{\Gamma'}^{(\rho)} = \sum_{\lambda \in J} \theta_{\Gamma}^{(\rho+\lambda)} \tag{164}$$

One can see that in our case $\Gamma' = \Gamma_+$ is the gluing of $\Gamma = \Gamma_-$ with the glue group

$$\mathrm{im}\, \upsilon_-|_{\ker \upsilon_+} \ \subset \ \Gamma_-^*/\Gamma_- \tag{165}$$

Since $b_2(B) = 0$ the only solutions of (142) are given by flat connections. The flat connections on B correspond to the elements of $T = H^2(B)$, while the flat connections on $\underset{\cdot}{B} = -M_3^- \sqcup M_3^+$ are in bijection with the elements of $T_- \oplus T_+$. In the case of an ordinary $U(1)$ gauge theory without Freed–Witten anomaly, the short exact sequence (160) determines which flat connections on the boundary can be extended to flat connections in the bulk B. Namely, a flat connection on the boundary given by $(\mu, \nu) \in H^2(\underset{\cdot}{B}) = T_- \oplus T_+$ originates from a flat connection in B if it is in the image of the map υ or, equivalently, in the kernel of ψ. The Vafa–Witten partition function of a cobordism $B \in \mathfrak{B}$ in this case is simply given by

$$Z_{\text{VW}}^{U(1)}[B]_{\mu,\nu} = \delta_{\psi(\mu,\nu)} \tag{166}$$

where

$$\delta_\lambda = \begin{cases} 1, & \lambda = 0 \\ 0, & \text{otherwise} \end{cases} \tag{167}$$

In the case when the $U(1)$ connection is replaced by the $U(1)$ part of the $Spin^c(4)$ connection one has to take into account the appropriate shift ψ_0:

$$Z_{\text{VW}}^{U(1)}[B]_{\mu,\nu} = \delta_{\psi(\mu,\nu)-\psi_0}. \tag{168}$$

In the abelian case the "gluing kernel" defined in Sect. 3.2 is trivial: $K^{\alpha\beta}[M_3] = \delta^{\alpha\beta}$ (therefore there is no difference between partition functions with upper and lower indices). Then we should have the following relation between the Vafa–Witten partition function on M_4^+, M_4^-, and B, cf. (92):

$$Z_{\text{VW}}^{U(1)}[M_4^+]_\nu \;=\; \sum_{\mu \in T_-} Z_{\text{VW}}^{U(1)}[B]_{\mu,\nu}\, Z_{\text{VW}}^{U(1)}[M_4^-]_\mu\,. \tag{169}$$

Since the abelian Vafa–Witten partition function on an arbitrary four-manifold of the form (13) is given by the theta function of the corresponding lattice (17), Eq. (169) can be viewed as the identity (164) that relates theta functions of the lattice Γ_- to the theta function of glued lattice Γ_+.

3.7 Composing Cobordisms

Now let us consider two four-manifolds $M_4^{(1)}$, $M_4^{(2)}$, both satisfying (124), such that $\partial M_4^{(1)} = M_3^a \sqcup M_3^b$ and $\partial M_4^{(2)} = M_3^b \sqcup M_3^c$. The 3-manifold M_3^b is supposed to be connected and have an opposite orientation in $M_4^{(1)}$ and $M_4^{(2)}$. The manifolds M_3^a and M_3^c can be empty. Then the new manifold $M_4^+ = M_4^{(1)} \cup M_4^{(2)}$ obtained by gluing $M_4^{(1)}$ and $M_4^{(2)}$ along M_3^b also has the properties (124). If we interpret $M_4^{(1)}$ as a cobordism between 3-manifolds M_3^b and M_3^a, and $M_4^{(2)}$ as a cobordism between M_3^c and M_3^b, then the resulting manifold M_4^+ is the composition of these two cobordisms. It is easy to see that this composition is a particular case of gluing described in the previous section. Namely, the manifold M_4^+ can be obtained by gluing $M_4^- = M_4^{(1)} \sqcup M_4^{(2)}$ with a basic cobordism, illustrated in Fig. 17,

$$B \;\cong\; M_3^a \times I \sqcup M_3^b \times I \sqcup M_3^c \times I \;\in\; \mathfrak{B} \tag{170}$$

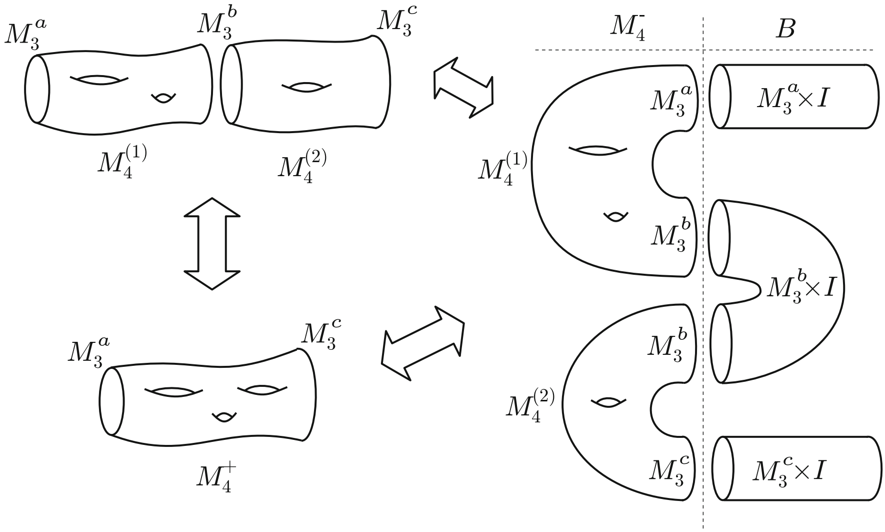


Fig. 17 Composition of cobordisms $M_4^{(1)} \circ M_4^{(2)} = M_4^+$ can be constructed by gluing $M_4^- = M_4^{(1)} \sqcup M_4^{(2)}$ with a basic cobordism $B \cong M_3^a \times I \sqcup M_3^b \times I \sqcup M_3^c \times I \in \mathfrak{B}$

where I is the interval. Let us denote $T^i = H^2(M_3^i)$, where $i = a, b, c$. Then, in the notations of the previous section, we have

$$\begin{aligned} T &= T^a \oplus T^b \oplus T^c \\ T_- &= T^a \oplus T^b \oplus T^b \oplus T^c \\ T_+ &= T^a \oplus T^c \end{aligned} \tag{171}$$

$$\upsilon_- : \ \lambda \oplus \mu \oplus \nu \ \longmapsto \ \lambda \oplus \mu \oplus (-\mu) \oplus \nu, \tag{172a}$$

$$\upsilon_+ : \ \lambda \oplus \mu \oplus \nu \ \longmapsto \ \lambda \oplus \nu.v \tag{172b}$$

As usual, let us denote $\Gamma_i \equiv H_2(M_4^{(i)})$ and $\Gamma_i^* \equiv H^2(M_4^{(i)})$. Then, the lattice Γ_+ is obtained by gluing of $\Gamma_1 \oplus \Gamma_2$ with the glue group

$$T^b \overset{\text{diag}}{\hookrightarrow} \Gamma_1^*/\Gamma_1 \oplus \Gamma_2^*/\Gamma_2 \cong (T^a \oplus T^b) \oplus (T^b \oplus T^c). \tag{173}$$

That is

$$\Gamma^+ = \left\{(\alpha + \mu, \beta - \mu) \mid \alpha \in \Gamma_1, \ \beta \in \Gamma_2, \ \mu \in T^b\right\}. \tag{174}$$

The Vafa–Witten partition functions of the manifolds $M_4^{(1)}$ and $M_4^{(2)}$ are given by:

$$Z_{\rm VW}^{U(1)}[M_4^{(1)}]_{\mu}^{\lambda}(q,x) \;=\; \sum_{\alpha\in\Gamma_1} q^{\frac{1}{2}\|\alpha+\lambda+\mu\|^2} x^{\alpha+\lambda+\mu}\,, \qquad (\lambda,\mu)\in T^a\oplus T^b\,, \tag{175a}$$

$$Z_{\rm VW}^{U(1)}[M_4^{(2)}]_{\nu}^{\mu}(q,y) \;=\; \sum_{\beta\in\Gamma_2} q^{\frac{1}{2}\|\beta-\mu+\mu_0+\nu\|^2} y^{\beta+\mu+\nu}\,, \qquad (\mu,\nu)\in T^b\oplus T^c\,, \tag{175b}$$

where the boundary condition μ on the boundary component M_3^b of $M_4^{(1)}$ is identified with the boundary condition $-\mu+\mu_0$ on $M_3^b\subset M_4^{(2)}$. The identity (169) in this case reads as:

$$\begin{aligned}
&\sum_{\mu} Z_{\rm VW}^{U(1)}[M_4^{(1)}]_{\mu}^{\lambda}(q,x)\; Z_{\rm VW}^{U(1)}[M_4^{(2)}]_{\nu}^{\mu}(q,y)\\
&\quad= \sum_{\alpha\in\Gamma_1,\,\beta\in\Gamma_2,\,\mu} q^{\frac{1}{2}\|\alpha+\lambda+\mu+\delta_1\|^2+\frac{1}{2}\|\beta-\mu+\mu_0+\nu+\Delta_2\|^2} x^{\alpha+\lambda+\mu+\Delta_1} y^{\beta-\mu+\mu_0+\nu+\Delta_2}\\
&\quad= \sum_{\gamma\in\Gamma_+} q^{\frac{1}{2}\|\gamma+(\lambda+\Delta_1)\oplus(\nu+\Delta_2+\mu_0)\|^2} (x,y)^{\gamma+(\lambda+\Delta_1)\oplus(\nu+\Delta_2+\mu_0)}\\
&\quad= Z_{\rm VW}^{U(1)}[M_4^{+}]_{\nu}^{\lambda}(q,(x,y))\,, \qquad (\lambda,\nu)\in T^a\oplus T^c\,.
\end{aligned} \tag{176}$$

so that the new shift due to the Freed–Witten anomaly is given by $\Delta=\Delta_1\oplus(\Delta_2+\mu_0)$.

3.8 Examples

Let us denote the 4-manifold associated with the Lie algebra $\mathfrak{g}$ of the *ADE* type as $M_4(\mathfrak{g})$ and the 4-manifold with the plumbing graph Υ by $M_4(\Upsilon)$, as in Sect. 2.1. For example,

$$A_{p-1} = M_4(\mathfrak{su}(p)) = M_4(\underbrace{\overset{-2}{\bullet}\!-\cdots-\!\overset{-2}{\bullet}}_{p-1}), \tag{177}$$

$$\begin{array}{c}\mathcal{O}(-p)\\ \downarrow\\ \mathbb{C}\mathbf{P}^1\end{array} \;=\; M_4(\overset{-p}{\bullet}), \tag{178}$$

$$\underbrace{\overline{\mathbb{C}\mathbf{P}}^2\#\ldots\#\overline{\mathbb{C}\mathbf{P}}^2}_{p}\setminus\{\mathrm{pt}\} \;=\; M_4(\underbrace{\overset{-1}{\bullet}\cdots\overset{-1}{\bullet}}_{p}). \tag{179}$$

As was previously mentioned, the lattice Γ for the 4-manifold $M_4(\mathfrak{g})$ coincides with the root lattice of $\mathfrak{g}$, while Γ^* is given by the corresponding weight lattice. The lattice Γ is always even and, therefore, $M_4(\mathfrak{g})$ is Spin and $\Delta=0$. Since level-1 characters are given by theta functions on the root lattice [KP], the formula (134) with $N = 1$,

$$Z_{\text{VW}}^{U(1)}[M_4(\mathfrak{g})]_\lambda \;=\; \chi_\lambda^{\hat{\mathfrak{g}}_1}\,, \tag{180}$$

also follows from (148). The abelian Vafa–Witten partition function of the A_p manifold was studied in detail in [DS08].

Let us point out that there is also the following relation between Vafa–Witten partition functions and affine characters:

$$Z_{\text{VW}}^{U(1)}[M_4(\overset{-p}{\bullet})]_\lambda(q,x) \;=\; \frac{1}{\eta(\tau)}\sum_{n\in\mathbb{Z}} q^{\frac{1}{2p}(pn+\lambda)^2} x^{pn+\lambda} \;\equiv\; \chi_\lambda^{\hat{\mathfrak{u}}(1)_p}\,, \qquad \lambda\in\mathbb{Z}_p \tag{181}$$

when p is even. This relation is a natural generalization of (180) since the one-dimensional lattice $H^2(M_4(\overset{-p}{\bullet}))$ can be interpreted as a weight lattice of $\hat{\mathfrak{u}}(1)_p$. Let us note that it is also consistent with the fact that $A_1 = M_4(\overset{-2}{\bullet})$ since

$$\chi_\lambda^{\widehat{\mathfrak{su}}(2)_1} = \chi_\lambda^{\hat{\mathfrak{u}}(1)_2}. \tag{182}$$

For general p one can write

$$Z_{\text{VW}}^{U(1)}[M_4(\overset{-p}{\bullet})]_\lambda(q,x) \;=\; \frac{1}{\eta(\tau)}\sum_{n\in\mathbb{Z}} q^{\frac{1}{2p}(pn+\lambda+\Delta)^2} x^{pn+\lambda+\Delta} \;\equiv\; \tilde{\chi}_\lambda^{\hat{\mathfrak{u}}(1)_p}\,, \qquad \lambda\in\mathbb{Z}_p \tag{183}$$

where $\Delta = 0$ if p is even and $\Delta = \frac{1}{2}$ if p is odd. Let us call $\tilde{\chi}^{\hat{\mathfrak{u}}(1)_p}$ the "twisted" $\hat{\mathfrak{u}}(1)_p$ character.

In Table 3 we present various examples of the gluing procedure described earlier. The corresponding gluings of lattices for many of these (and other) examples can be found in [GL92, GL92]. Let us note that in Example 3 one can choose the gluing cobordism to be a cylinder with a hole $B = S^3/\mathbb{Z}_p \times I \setminus \text{pt}$, i.e., one can just glue two components of M_4^- along their boundaries (and then cut out a hole) in order to obtain M_4^+. In Examples 8, 9 the cobordism B is homologically equivalent to a cylinder with a hole, but not topologically, since the boundaries of E_{8-n} and A_n are only homologically equivalent. Consider Example 2 in some detail. In general it is not possible to glue $M_4(\overset{-k}{\bullet})$ with $M_4(\overset{-k}{\bullet})$, because although the boundaries are the same, they do not have opposite orientations. However, when $k = p^2 + 1$ for some integer p there exists an orientation *reversing* diffeomorphism φ of $L(k, 1)$ such that

$$\begin{aligned} \varphi^* : H^2(L(k,1)) &\longrightarrow H^2(L(k,1)) \cong \mathbb{Z}_k \\ \rho &\longmapsto p\rho \end{aligned} \tag{184}$$

Table 3 Examples of gluing $M_4^- \overset{B}{\rightsquigarrow} M_4^+$

	Original 4-manifold M_4^-	End result M_4^+	Homological data of $B \in \mathfrak{B}$ ($b_2(B) = 0$)	
	$T_- = H^2(M_4^-)$	$T_+ = H^2(M_4^+)$	$T = H^2(B)$	$\upsilon : T \to T_- \oplus T_+$, $\psi : T_- \oplus T_+ \to T$
1	$M_4(\overset{-p^2}{\bullet})$ $T_- = \mathbb{Z}_{p^2}$	$M_4(\overset{-1}{\bullet})$ $T_+ = 0$	$\mathbb{Z}_p$	$\upsilon(\rho) = p\rho$ $\psi(\mu) = (\mu \mod p)$
2	$M_4(\overset{-p^2-1}{\bullet}) \sqcup$ $M_4(\overset{-p^2-1}{\bullet})$ $T_- = \mathbb{Z}_{p^2+1}$	$M_4(\overset{-1}{\bullet} \overset{-1}{\bullet})$ $T_+ = 0$	$\mathbb{Z}_{p^2+1}$	$\upsilon(\rho) = \rho \oplus p\rho$ $\psi(\mu \oplus \nu) = (p\mu - \nu)$
3	$A_{p-1} \sqcup M_4(\overset{-p}{\bullet})$ $T_- = \mathbb{Z}_p \oplus \mathbb{Z}_p$	$M_4(\underbrace{\overset{-1}{\bullet} \cdots \overset{-1}{\bullet}}_{p})$ $T_+ = 0$	$\mathbb{Z}_p$	$\upsilon(\rho) = \rho \oplus \rho$ $\psi(\mu \oplus \nu) = (\mu - \nu)$
4	$A_{p-1} \sqcup M_4(\overset{-p(p+1)}{\bullet})$ $T_- = \mathbb{Z}_p \oplus \mathbb{Z}_p \oplus \mathbb{Z}_{p+1}$	A_p $T_+ = \mathbb{Z}_{p+1}$	$\mathbb{Z}_p \oplus \mathbb{Z}_{p+1}$	$\upsilon(\rho \oplus \lambda) = \rho \oplus \rho \oplus \lambda \oplus \lambda$ $\psi(\mu \oplus \nu \oplus \rho \oplus \lambda) = (\mu - \nu) \oplus (\rho - \lambda)$
5	$M_4(\overset{-a_1}{\bullet}\!-\cdots-\!\overset{-a_n}{\bullet}) \sqcup$ $M_4(\overset{-p_np_{n+1}}{\bullet})$ where $p_{n+1} = a_np_n - p_{n-1}$ $T_- = \mathbb{Z}_{p_n} \oplus \mathbb{Z}_{p_n} \oplus \mathbb{Z}_{p_{n+1}}$	$M_4(\overset{-a_1}{\bullet}\!-\cdots-\!\overset{-a_{n+1}}{\bullet})$ $T_+ = \mathbb{Z}_{p_{n+1}}$	$\mathbb{Z}_{p_n} \oplus \mathbb{Z}_{p_{n+1}}$	$\upsilon(\rho \oplus \lambda) = \rho \oplus \rho \oplus \lambda \oplus \lambda$ $\psi(\mu \oplus \nu \oplus \rho \oplus \lambda) = (\mu - \nu) \oplus (\rho - \lambda)$
6	$A_3 \sqcup M_4(\overset{-4}{\bullet})$ $T_- = \mathbb{Z}_4 \oplus \mathbb{Z}_4$	D_4 $T_+ = \mathbb{Z}_2 \oplus \mathbb{Z}_2$	$\mathbb{Z}_4 \oplus \mathbb{Z}_2$	$\upsilon(\mu \oplus \nu) = \mu \oplus (\mu + 2\nu) \oplus (\mu \mod 2) \oplus \nu$ $\psi(\mu \oplus \nu \oplus \rho \oplus \lambda) = (\nu - \mu - 2\lambda) \oplus ((\mu \mod 2) - \rho)$
7	D_8 $T_- = \mathbb{Z}_2 \oplus \mathbb{Z}_2$	E_8 $T_+ = 0$	$\mathbb{Z}_2$	$\upsilon(\rho) = \rho \oplus 0$ $\psi(\mu \oplus \nu) = \nu$
8	$E_7 \sqcup A_1$ $T_- = \mathbb{Z}_2 \oplus \mathbb{Z}_2$	E_8 $T_+ = 0$	$\mathbb{Z}_2$	$\upsilon(\rho) = \rho \oplus \rho$ $\psi(\mu \oplus \nu) = (\mu - \nu)$
9	$E_6 \sqcup A_2$ $T_- = \mathbb{Z}_3 \oplus \mathbb{Z}_3$	E_8 $T_+ = 0$	$\mathbb{Z}_3$	$\upsilon(\rho) = \rho \oplus \rho$ $\psi(\mu \oplus \nu) = (\mu - \nu)$
10	A_8 $T_- = \mathbb{Z}_9$	E_8 $T_+ = 0$	$\mathbb{Z}_3$	$\upsilon(\rho) = 3\rho$ $\psi(\mu) = (\mu \mod 3)$
11	$A_4 \sqcup A_4$ $T_- = \mathbb{Z}_5 \oplus \mathbb{Z}_5$	E_8 $T_+ = 0$	$\mathbb{Z}_5$	$\upsilon(\rho) = \rho \oplus 2\rho$ $\psi(\mu \oplus \nu) = (2\mu - \nu)$

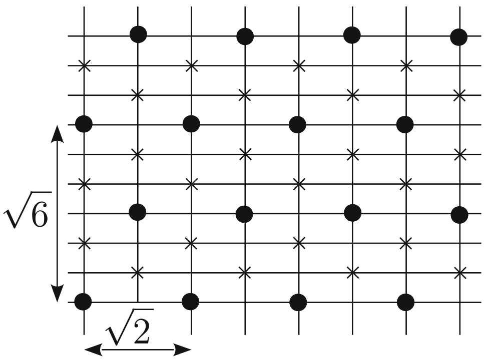


Fig. 18 Gluing of A_1 and $M_4(\overset{-6}{\bullet})$ gives A_2

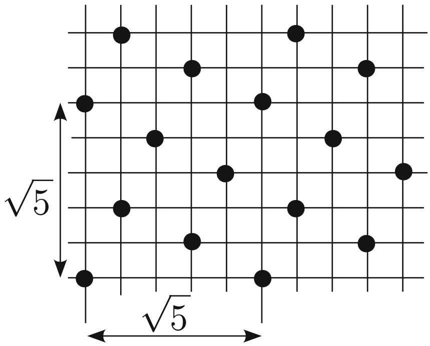


Fig. 19 Gluing of $M_4(\overset{-5}{\bullet})$ and $M_4(\overset{-5}{\bullet})$ gives $M_4(\overset{-1}{\bullet}\overset{-1}{\bullet})$

It is an automorphism of $\mathbb{Z}_k$ because p and $k = p^2+1$ are coprime. One can also glue A_{p^2} with A_{p^2} using the same prescription (cf. Example 11). The gluings of lattices in Examples 2 and 3 are illustrated in Figs. 18 and 19.

Let us consider in some detail the gluing in Example 3 when p is even. This example is rather interesting because both of the original 4-manifolds A_{p-1} and $M_4(\overset{-p}{\bullet})$ are Spin, but the resulting 4-manifold $M_4(\overset{-1}{\bullet} \dots \overset{-1}{\bullet})$ is not Spin (since the corresponding lattice $\mathbb{Z}^p$ is not even). What is going on here? The explanation is very instructive and reveals new aspects of the Freed–Witten anomaly in the presence of boundaries.

Each of the original "pieces", A_{p-1} and $M_4(\overset{-p}{\bullet})$, admits a unique Spin structure. However, the restrictions of these Spin structures to the boundary 3-manifold M_3, along which one must glue these pieces in order to produce $M_4(\overset{-1}{\bullet} \dots \overset{-1}{\bullet})$, are different. To be a little more precise, as in (92) consider the gluing map between the boundaries:

$$\varphi : \underset{\cdot}{\mathrm{A}}_{p-1} \to \underset{\cdot}{\mathrm{M}}_4(\overset{-p}{\bullet}) \tag{185}$$

If we introduce Spin structures on A_{p-1} and $M_4(\overset{-p}{\bullet})$, the map φ does not lift to a map between the restrictions of the Spin structures on the boundaries. This is why it is not possible to construct a Spin structure on $M_4(\overset{-1}{\bullet} \dots \overset{-1}{\bullet})$ from the Spin structures on A_{p-1} and $M_4(\overset{-p}{\bullet})$.

Nevertheless, it is possible to lift φ to a map between the restrictions of Spin^c structures on A_{p-1} and $M_4(\overset{-p}{\bullet})$. Since $Spin(4)$ holonomies on the boundaries do not match, the holonomies of the $U(1)$ part of $Spin^c(4)$ should be identified with -1 factor which corresponds to the shift by $\frac{p}{2}$ in the $\mathbb{Z}_p$ space of flat connections on the boundaries. One can check that indeed

$$\begin{aligned}
&\sum_{\lambda\in\mathbb{Z}_p} Z_{\text{VW}}^{U(1)}[M_4(\overset{-p}{\bullet})]_{\lambda+p/2}(q,x^{\perp})\, Z_{\text{VW}}^{U(1)}[M_4(\overset{-p}{\bullet})]^{\lambda}(q,x^{\parallel}) \\
&\quad = \sum_{\lambda\in\mathbb{Z}_p} \chi_{\lambda+p/2}^{\hat{\mathfrak{u}}(1)_p}(q,x^{\perp})\, \chi_{\lambda}^{\widehat{\mathfrak{su}}(p)_1}(q,x^{\parallel}) \;=\; \tilde{\chi}^{\hat{\mathfrak{u}}(p)_1}(q,x) \;\equiv\; \prod_{i=1}^{p} \tilde{\chi}^{\hat{\mathfrak{u}}(1)_1}(q,x_i) \\
&\quad = Z_{\text{VW}}^{U(1)}[M_4(\underbrace{\overset{-1}{\bullet}\cdots\overset{-1}{\bullet}}_{p})]
\end{aligned} \tag{186}$$

where the splitting of parameters $x = (x^{\perp}, x^{\parallel})$ is such that $x^{\perp} = (\prod_i x_i)^{1/p}$. A version of this relation without shifts due to Freed–Witten anomaly was considered in [DHSV07, DS08].

In general, a gluing of the form

$$M_4(\mathfrak{g}^{(1)}) \sqcup \ldots \sqcup M_4(\mathfrak{g}^{(n)}) \sqcup M_4(\overset{-p_1}{\bullet}) \sqcup \ldots \sqcup M_4(\overset{-p_m}{\bullet}) \quad \overset{\mathcal{B}}{\rightsquigarrow} \quad M_4(\mathfrak{g}) \tag{187}$$

where all p_i are even, $\mathfrak{g}^{(j)}$ and $\mathfrak{g}$ are of ADE type, corresponds to the embedding of the associated algebras:

$$\mathfrak{g}_1^{(1)} \oplus \ldots \oplus \mathfrak{g}_1^{(n)} \oplus \mathfrak{u}(1)_{p_1} \oplus \ldots \oplus \mathfrak{u}(1)_{p_m} \;\subset\; \mathfrak{g} \tag{188}$$

where the subscripts denote the indices of the embeddings.

Let us recall that the index ℓ of the embedding $\mathfrak{k}_\ell \subset \mathfrak{g}$ is defined as the ratio between the normalized Killing form of $\mathfrak{g}$ restricted to the subspace $\mathfrak{k}$ and the normalized Killing form of $\mathfrak{k}$. In other words, the root lattice of $\mathfrak{k}$ is rescaled by the factor of $\sqrt{\ell}$ when embedded into the root lattice of $\mathfrak{g}$. For the corresponding affine Lie algebras, representations of $\hat{\mathfrak{g}}$ at level k decompose into representations of $\hat{\mathfrak{k}}$ at level ℓk:

$$\chi_{\lambda}^{\hat{\mathfrak{g}}_k} \;=\; \sum_{\mu} b_{\lambda}^{\mu}\, \chi_{\mu}^{\hat{\mathfrak{k}}_{\ell k}}\,. \tag{189}$$

The coefficients b_{λ}^{μ} are called branching functions of the embedding $\mathfrak{k}_\ell \subset \mathfrak{g}$.

If $B \in \mathfrak{B}$, that is $b_2(B) = 0$, the total rank on both sides of (188) is the same:

$$\sum_{i=1}^{n} \operatorname{rank} \mathfrak{g}^{(i)} + m = \operatorname{rank} \mathfrak{g}. \tag{190}$$

Then, taking into account (180) and (183), the identity (169) can be interpreted as a decomposition of the characters:

$$\boxed{\chi_{\lambda}^{\hat{\mathfrak{g}}_1} \;=\; \sum_{\mu,\rho} Z_{\rm VW}^{U(1)}[B]_{\lambda}^{\mu_1\ldots\mu_n\rho_1\ldots\rho_m}\;\chi_{\mu_1}^{\hat{\mathfrak{g}}_1^{(1)}}\cdots\chi_{\mu_n}^{\hat{\mathfrak{g}}_1^{(n)}}\;\chi_{\rho_1}^{\hat{\mathfrak{u}}(1)_{p_1}}\cdots\chi_{\rho_m}^{\hat{\mathfrak{u}}(1)_{p_m}}} \tag{191}$$

so that the Vafa–Witten partition function of B plays the role of branching functions for the embedding (188) at level 1. As was shown earlier, the abelian Vafa–Witten partition function of $B \in \mathfrak{B}$ does not depend on τ. This corresponds to the fact that the embedding (188) is always conformal at level 1.

Now let us define $\widetilde{B}$ as B glued with $M_4(\overset{-p_1}{\bullet}) \sqcup \ldots \sqcup M_4(\overset{-p_m}{\bullet})$ along the common boundary components. This 4-manifold $\widetilde{B}$ is no longer in $\mathfrak{B}$ and has $b_2(\widetilde{B}) = m$. It can be considered as a cobordism for the following gluing:

$$M_4(\mathfrak{g}^{(1)}) \sqcup \ldots \sqcup M_4(\mathfrak{g}^{(n)}) \quad \overset{\widetilde{B}}{\rightsquigarrow} \quad M_4(\mathfrak{g})\,. \tag{192}$$

The identity (191) can be rewritten as

$$\chi_{\lambda}^{\hat{\mathfrak{g}}_1} \;=\; \sum_{\mu} Z_{\rm VW}^{U(1)}\,[\widetilde{B}]_{\lambda}^{\mu_1\ldots\mu_n}\;\chi_{\mu_1}^{\hat{\mathfrak{g}}_1^{(1)}}\cdots\chi_{\mu_n}^{\hat{\mathfrak{g}}_1^{(n)}} \tag{193}$$

and, therefore, $Z_{\rm VW}^{U(1)}[\widetilde{B}]$ plays the role of the level-1 branching functions for the embedding

$$\mathfrak{g}_1^{(1)} \oplus \ldots \oplus \mathfrak{g}_1^{(n)} \;\subset\; \mathfrak{g} \tag{194}$$

where all Lie algebras are of *ADE* type.

3.9 Non-abelian Generalizations

As was already mentioned in Sect. 3.3, the non-abelian generalization of (180) is given by

$$Z_{\rm VW}^{U(N)}[M_4(\mathfrak{g})]_{\rho} \;=\; \chi_{\rho}^{\hat{\mathfrak{g}}_N} \tag{195}$$

Hence, the Vafa–Witten partition function of a cobordism $\widetilde{B}$ in (192) should coincide with the branching functions for the embedding (194) at level N:

$$\boxed{Z_{\rm VW}^{U(N)}\,[\widetilde{B}]_{\lambda}^{\mu_1\ldots\mu_n} \;=\; \text{branching function } b_{\lambda}^{\mu_1\ldots\mu_n}}$$

Since the lattice $H^2(M_4(\overset{-p}{\bullet}))$ is one-dimensional it is natural to expect that the corresponding Vafa–Witten partition function can be expressed in terms of $\hat{\mathfrak{u}}(1)$ characters. As a non-abelian generalization of (183) one can propose that

$$Z_{\text{VW}}^{U(N)}[M_4(\overset{-p}{\bullet})]_\lambda(q,x) \;=\; \sum_\mu C_\lambda^\mu(q)\,\tilde{\chi}_\mu^{\hat{\mathfrak{u}}(1)_{pN}}(q,x) \tag{196}$$

with some coefficients C_λ^μ independent of x. This is consistent with the fact that $M_4(\overset{-2}{\bullet}) = A_1$ because the characters of $\widehat{\mathfrak{su}}(2)$ can be decomposed in terms of the $\hat{\mathfrak{u}}(1)$ characters, where $\mathfrak{u}(1)$ is embedded as a Cartan subalgebra of $\mathfrak{su}(2)$ with index 2:

$$Z_{\text{VW}}^{U(N)}[M_4(\overset{-2}{\bullet})]_\lambda(q,x) = Z_{\text{VW}}^{U(N)}[A_1]_\lambda(q,x) = \chi_\lambda^{\widehat{\mathfrak{su}}(2)_N}(q,x) = \sum_\mu C_\lambda^\mu(q)\chi_\mu^{\hat{\mathfrak{u}}(1)_{2N}}(q,x) \tag{197}$$

Hence, in this case C_λ^μ are the branching functions for the embedding $\mathfrak{u}(1)_2 \subset \mathfrak{su}(2)$. The formula (196) is also in agreement with the results of [AOSV05].

From (196) and (195) it follows that $Z_{\text{VW}}^{U(N)}[B]$ for the cobordism B in (187) is given, up to coefficients C, by level-N characters of the coset for the embedding (188):

$$\frac{G}{G^{(1)} \times \ldots \times G^{(n)} \times \underbrace{U(1) \times \ldots \times U(1)}_{m}}. \tag{198}$$

Note, such coset spaces are Kähler manifolds because of the property (190). This suggests that the corresponding 2d theories $T[B]$ may have a realization in terms of $(0,2)$ gauged WZW theories studied in [Joh95, BJKZ96].

Now let us discuss various consequences and consistency checks of the proposed relation between cobordisms and branching functions. In [VW94] it was argued that under the blow up of M_4 (that is taking the connected sum with $\overline{\mathbb{C}\mathbf{P}}^2$) the $SU(N)$ partition function on M_4 is multiplied by the character of $\widehat{\mathfrak{su}}(N)_1$:

$$Z_{\text{VW}}^{SU(N)}[M_4\#\overline{\mathbb{C}\mathbf{P}}^2] \;=\; Z_{\text{VW}}^{SU(N)}[M_4]\,\chi^{\widehat{\mathfrak{su}}(N)_1}\,. \tag{199}$$

Based on our experience with abelian theory discussed in the previous section, it is then natural to propose the following generalization to the case of $U(N)$ gauge group and non-compact 4-manifolds:

$$Z_{\text{VW}}^{U(N)}\left[M_4 \natural \left(\overline{\mathbb{C}\mathbf{P}}^2 \setminus \{\text{pt}\}\right)\right](\tau,x) \;=\; Z_{\text{VW}}^{U(N)}[M_4](\tau,x^{\|})\,\tilde{\chi}^{\hat{\mathfrak{u}}(N)_1}(\tau,x^{\perp}) \tag{200}$$

where $\natural$ denotes the boundary connected sum, $x = (x^{\|}, x_{\perp})$, $x^{\|} \in \exp(H_2(M_4) \otimes \mathbb{C})$, and $x^{\perp} \in \exp(H_2(\overline{\mathbb{C}\mathbf{P}}^2 \setminus \{\text{pt}\}) \otimes \mathbb{C}) \cong \mathbb{C}^*$. The "twisted" $\hat{\mathfrak{u}}(N)_1$ character $\tilde{\chi}^{\hat{\mathfrak{u}}(N)_1}$

is defined as in (186). The parameter $x \in \mathbb{C}^*$ plays the role of the coordinate along the diagonal $\mathfrak{u}(1)$ of $\mathfrak{u}(N)$, and the coordinates in the other directions of the Cartan subalgebra are set to zero. If the manifold M_4 is constructed by the plumbing graph Υ, the relation (200) looks like

$$Z_{\text{VW}}^{U(N)}[M_4(\Upsilon \sqcup \overset{-1}{\bullet})] = Z_{\text{VW}}^{U(N)}[M_4(\Upsilon)]\, \tilde{\chi}^{\hat{\mathfrak{u}}(N)_1}. \tag{201}$$

In particular:

$$Z_{\text{VW}}^{U(N)}[M_4(\underbrace{\overset{-1}{\bullet} \cdots \overset{-1}{\bullet}}_{p})] = \prod_{i=1}^{p} \tilde{\chi}^{\hat{\mathfrak{u}}(N)_1}(q, x_i). \tag{202}$$

Let us note that the "twisted" $\hat{\mathfrak{u}}(N)_1$ character is given by the product of N standard theta functions with odd characteristics:

$$\tilde{\chi}^{\hat{\mathfrak{u}}(N)_1}(q, z) = \prod_{j=1}^{N} \frac{1}{\eta(q)} \sum_{n_j \in \mathbb{Z}} q^{\frac{(n_j+1/2)^2}{2}} z^{n_j+1/2} \equiv \prod_{j=1}^{N} \frac{\theta_2(q, z_j)}{\eta(q)}. \tag{203}$$

Therefore, (202) can be rewritten as

$$Z_{\text{VW}}^{U(N)}[M_4(\underbrace{\overset{-1}{\bullet} \cdots \overset{-1}{\bullet}}_{p})](q, x) = \prod_{i=1}^{p}\prod_{j=1}^{N} \frac{\theta_2(q, x_i)}{\eta(q)} = \tilde{\chi}^{\hat{\mathfrak{u}}(Np)_1}(q, x) \tag{204}$$

where the components x_i play the role of the coordinates in the diagonal directions of p copies of the $\mathfrak{u}(N)$ subalgebra in $\mathfrak{u}(Np)$. In [DHSV07] it was shown that the embedding (which is conformal at level 1)

$$\mathfrak{su}(N)_p \oplus \mathfrak{u}(1)_{pN} \oplus \mathfrak{su}(p)_N \subset \mathfrak{u}(Np), \tag{205}$$

leads to the following relation between the "untwisted" characters:

$$\begin{aligned} \prod_{i=1}^{p}\prod_{j=1}^{N} \frac{\theta_3(q, x_i y_j)}{\eta(q)} &\equiv \chi^{\hat{\mathfrak{u}}(Np)_1}(q, \{x, y\}) \\ &= \sum_{[\lambda]} \sum_{a=1}^{N} \sum_{b=1}^{p} \chi_{\sigma_N^a(\lambda)}^{\widehat{\mathfrak{su}}(N)_p}(q, y^{\parallel})\, \chi_{|\lambda|+ap+bN}^{\hat{\mathfrak{u}}(1)_{Np}}(x^{\perp} y^{\perp})\, \chi_{\sigma_p^b(\lambda^t)}^{\widehat{\mathfrak{su}}(p)_N}(q, x^{\parallel}) \end{aligned} \tag{206}$$

where $x^{\perp} = (\prod_i x_i)^N$, $y^{\perp} = (\prod_j x_j)^p$, σ_N and σ_p denote the generators of outer automorphisms groups $\mathbb{Z}_N$ and $\mathbb{Z}_p$ of $\widehat{\mathfrak{su}}(N)$ and $\widehat{\mathfrak{su}}(p)$, respectively, λ denotes an integrable representation of $\widehat{\mathfrak{su}}(p)_N$ associated with a Young diagram, and λ^t

denotes an integrable representation of $\widehat{\mathfrak{su}}(N)_p$ associate to the transposed Young diagram. The first sum on the right-hand side of this expression is performed over the orbits $[\lambda]$ of λ with respect to the action of the outer automorphism group. Finally, $|\lambda|$ stands for the number of boxes in the Young diagram associated with λ. See [DHSV07] for the details.

When $p = 1$ and $y = 0$, it follows from (206) that

$$Z_{\rm VW}^{U(N)}[M_4(\overset{-1}{\bullet})] \;=\; \chi^{\hat{\mathfrak{u}}(N)_1}(q,x) \;=\; \sum_{\lambda} \chi_\lambda^{\widehat{\mathfrak{su}}(N)_1}(q,0)\, \chi_\lambda^{\hat{\mathfrak{u}}(1)_N}(q,x) \tag{207}$$

and, therefore, the coefficients C in (196) in the case $p = 1$ are given by the characters of $\widehat{\mathfrak{su}}(N)_1$.

Now let us consider the Example 3 from Table 3:

$$A_{p-1} \sqcup M_4(\overset{-p}{\bullet}) \overset{B}{\rightsquigarrow} M_4(\underbrace{\overset{-1}{\bullet}\cdots\overset{-1}{\bullet}}_{p}). \tag{208}$$

As was mentioned earlier, B is topologically a cylinder with a hole: $B \cong L(p,1) \times I \setminus \{\mathrm{pt}\}$. One can expect the following identify for the corresponding non-abelian Vafa–Witten partition functions:

$$\begin{aligned} &Z_{\rm VW}^{U(N)}[M_4(\underbrace{\overset{-1}{\bullet}\cdots\overset{-1}{\bullet}}_{p})](q,x) \\ &\quad= \sum_{\lambda,\mu} Z_{\rm VW}^{U(N)}[M_4(\overset{-p}{\bullet})]_\lambda(q,x^{\perp})\, Z_{\rm VW}^{U(N)}[B]^{\lambda,\mu}(q)\, Z_{\rm VW}^{U(N)}[A_{p-1}]_\mu(q,x^{\parallel})\,. \end{aligned} \tag{209}$$

Taking into account

$$Z_{\rm VW}^{U(N)}[A_{p-1}]_\mu(q,x^{\parallel}) \;=\; \chi_\mu^{\widehat{\mathfrak{su}}(p)_N}(q,x^{\parallel}) \tag{210}$$

combined with (202) and (196), one can interpret (209) as the "twisted" version of the identity (206) in the case where y is set to zero.

3.10 Linear Plumbings and Quiver Structure

From Example 5 in Table 3 it follows that one can build the plumbing $\overset{a_1}{\bullet}\!-\cdots-\!\overset{a_n}{\bullet}$ step by step, attaching one node at a time. Moreover, as we explained in Sect. 2.2, the boundary 3-manifold is the Lens space, $M_3(\overset{a_1}{\bullet}\!-\cdots-\!\overset{a_n}{\bullet}) = L(p_n,q_n)$, where p_n/q_n is given by the continued fraction (43) associated with the string of integers $(a_1,\ldots,a_n)$. Therefore, the gluing discussed in Sects. 2.14 and 3.4

$$M_4(\overset{a_1}{\bullet}\!-\cdots-\!\overset{a_n}{\bullet}) \quad \rightsquigarrow \quad M_4(\overset{a_1}{\bullet}\!-\cdots-\!\overset{a_n}{\bullet}\!\!-\!\!\!-\!\!\overset{a_{n+1}}{\bullet}) \tag{211}$$

can be achieved with a certain cobordism $B^{p_{n+1},q_{n+1}}_{p_n,q_n}$ from the family (124), which is uniquely determined by the properties

$$\begin{aligned} \partial B^{p_{n+1},q_{n+1}}_{p_n,q_n} &= -L(p_n,q_n) \sqcup L(p_{n+1},q_{n+1}) \\ b_2(B^{p_{n+1},q_{n+1}}_{p_n,q_n}) &= 1 \end{aligned} \tag{212}$$

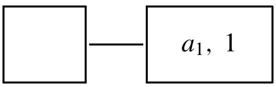
The cobordism $B^{p_{n+1},q_{n+1}}_{p_n,q_n}$ can be obtained by joining the cobordism B in Example 5 of Table 3 with $M_4(\overset{-p_n p_{n+1}}{\bullet})$. Let us note that the Lens spaces $L(p,q)$ are homologically equivalent for different values of q and have $H_1(L(p,q)) = \mathbb{Z}_p$. A manifestation of this fact is that the abelian Vafa–Witten partition function of the cobordism $B^{p',q'}_{p,q}$ depends only on p and p', and is given by

$$Z^{U(1)}_{\text{VW}}[B^{p',q'}_{p,q}]^{j'}_{j} = \sum_{n\in\mathbb{Z}} q^{\frac{pp'}{2}\left(n-\frac{j}{p}+\frac{j'}{p'}\right)^2} x^{pp'n-p'j+pj'}, \qquad j\in\mathbb{Z}_p,\ j'\in\mathbb{Z}_{p'} \tag{213}$$

when p and p' are even.

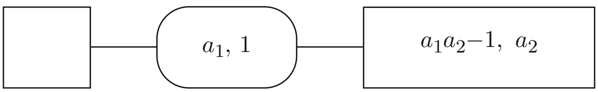
This gluing procedure can be formally encoded in a quiver diagram where every vertex is labeled by pair of integers. This quiver can be interpreted as a quiver description of the corresponding 2d theory $T[M_4]$. A four-manifold with $L(p,q)$ boundary has a "flavor symmetry vertex" $\boxed{p,q}$. When the cobordism $B^{p',q'}_{p,q}$ is glued to it to produce the $L(p',q')$ boundary, we "gauge" the $\boxed{p,q}$ vertex with the $\boxed{p,q}$ vertex of the "bifundamental" $\boxed{p,q}-\boxed{p',q'}$.

Let us illustrate this gluing procedure with an example. Consider the plumbing $\overset{a_1}{\bullet}\!-\!\!\!-\!\overset{a_2}{\bullet}$. We start with the node $\overset{a_1}{\bullet}$. The corresponding manifold $M_4(\overset{a_1}{\bullet})$ can be considered as a cobordism from the empty space to $L(a_1,1)$. Therefore, the quiver associated with it looks like



(214)

The boundary of the space after adding the plumbing node $\overset{a_2}{\bullet}$ is another Lens space $L(a_1a_2-1,a_2)$. This space is obtained by gluing $M_4(\overset{a_1}{\bullet})$ with $B^{a_1,1}_{a_1a_2-1,a_2}$. After "gauging" the node $\boxed{a_1,1}$ we get the quiver



(215)

Clearly, the associated quiver in general depends on the plumbing sequence. We expect each quiver to give a 2d $\mathcal{N} = (0, 2)$ theory and theories associated with the same plumbing to be dual to each other. For the purposes of computing Z_{VW}, the "flavor symmetry node" stands for a boundary condition label. "Gauging" this node means summing over all such labels.

Let us consider in more detail how this works in the case when all $a_i = -2$. The 4-manifold constructed by the plumbing with n nodes is then A_n, and adding one extra node (cf. Example 4 in Table 3) can be realized by the cobordism $B^{n+2,n+1}_{n+1,n}$. As was explained in Sect. 3.3, the relevant ingredients have the form:

$$Z^{U(N)}_{\text{VW}}[A_{n+1}]_\rho(q,x) \;=\; \sum_\lambda Z^{U(N)}_{\text{VW}}[B^{n+2,n+1}_{n+1,n}]^\lambda_\rho(q,x^\perp)\, Z^{U(N)}_{\text{VW}}[A_n]_\lambda(q,x^\parallel)\,, \tag{216}$$

$$Z^{U(N)}_{\text{VW}}[A_n]_\lambda \;=\; \chi^{\widehat{\mathfrak{su}}(n+1)_N}_\lambda\,, \tag{217}$$

$$Z^{U(N)}_{\text{VW}}[B^{n+2,n+1}_{n+1,n}]^\lambda_\rho \;=\; \chi^{\widehat{\mathfrak{su}}(n+2)_N/\widehat{\mathfrak{su}}(n+1)_N}_{\lambda,\rho}\,. \tag{218}$$

This suggests that $T[B^{n+2,n+1}_{n+1,n}]$ may have a realization in terms of $\widehat{\mathfrak{su}}(n+2)_N/\widehat{\mathfrak{su}}(n+1)_N$ coset WZW. Direct realization in terms of $(0,2)$ WZW models considered in [Joh95, BJKZ96] is difficult because the coset space does not have a complex structure. However, as we will show below, it is easy to interpret the Vafa–Witten partition function on $B^{n+2,n+1}_{n+1,n}$ if we make a certain transformation changing *discrete* labels associated with boundary conditions to *continuous* variables. This transformation can be interpreted as a change of basis in TQFT Hilbert spaces associated with boundaries. Namely, let us define the Vafa–Witten partition function on A_n in the continuous basis as

$$Z^{U(N)}_{\text{VW}}[A_{n-1}](q,x|z) \;:=\; \sum_\rho \chi^{\hat{\mathfrak{u}}(N)_n}_{\tilde{\rho}}(q,z)\, Z^{U(N)}_{\text{VW}}[A_{n-1}]_\rho(q,x) \tag{219}$$

where we used that, due to the level-rank duality, there is a one-to-one correspondence $\rho \leftrightarrow \tilde{\rho}$ between integrable representations of $\widehat{\mathfrak{su}}(n)_N$ and $\hat{\mathfrak{u}}(N)_n$ realized by transposing the corresponding Young diagrams. Namely,

$$\chi^{\hat{\mathfrak{u}}(N)_n}_{\tilde{\rho}}(q,z) \;=\; \sum_{a=1}^N \chi^{\hat{\mathfrak{u}}(1)_{Nn}}_{|\rho|+an}(q,z^\perp)\, \chi^{\widehat{\mathfrak{su}}(N)_n}_{\sigma^a_N(\rho^t)}(q,z^\parallel) \tag{220}$$

in the notations of the formula (206).

The fugacities z in (219) can be interpreted as fugacities for flavor symmetry of $T[M_4]$ associated with the boundary $M_3 = \partial M_4$. This symmetry is the gauge symmetry of $T[M_3]$. Gluing two 4-manifolds with along the common boundary M_3 corresponds to integrating over z, that is gauging the common flavor symmetry associated with z. Naively, the fugacities x have different nature since they are associated with 2-cycles, not three-dimensional boundaries. However, one can expect a relation between them since one can always produce a three-dimensional boundary by excising a tabular neighborhood of a 2-cycle.

It is convenient to introduce the q-theta function defined as:

$$\theta(w;q) \ := \ \prod_{r=0}^{\infty}(1-q^r w)(1-q^{r+1}/w) \ = \ (w;q)_\infty (q/w;q)_\infty \tag{221}$$

where

$$(w;q)_s \ := \ \prod_{r=0}^{s-1}(1-wq^r) \tag{222}$$

is the q-Pochhammer symbol. From (206) it follows then that in the continuous basis the Vafa–Witten partition function takes a remarkably simple form:

$$Z_{\text{VW}}^{U(N)}[A_{n-1}](q,x|z) \ = \ q^{-\frac{nN}{24}} \prod_{i=1}^{n}\prod_{j=1}^{N} \theta(-q^{\frac{1}{2}} x_i z_j; q) \tag{223}$$

where the fugacities x are represented by $x_i \in \mathbb{C}^*$, $i = 1 \dots n$ satisfying $\prod_{i=1}^{n} x_i = 1$.

Now, in the *continuous basis*, the right-hand side of (223) can be interpreted as the flavored elliptic genus (9) of nN Fermi multiplets, possibly with a superpotential (to account for the q shift in the argument). In [DHSV07] the transition from the $\hat{\mathfrak{u}}(Nn)_1$ character in the right-hand side of (223) to the $\widehat{\mathfrak{su}}(n)_N$ character in the right-hand side of (217) was interpreted as gauging degrees of freedom of D4-branes obtained by a compactification of M5-branes.

As we show explicitly in Appendix 2 for $N = 2$ and conjecture for general N, the characters satisfy the following orthogonality condition:

$$\oint \frac{dz}{2\pi i z}\, \mathcal{I}_V^{U(N)}(q,z)\, \chi_\lambda^{\hat{\mathfrak{u}}(N)_n}(q,z)\, \chi_{\lambda'}^{\hat{\mathfrak{u}}(N)_n}(q,z) \ = \ C_\lambda(q)\delta_{\lambda,\lambda'} \tag{224}$$

where

$$\mathcal{I}_V^{U(N)}(q,z) \ = \ (q;q)_\infty^{2N} \prod_{i \neq j} \theta(z_i/z_j;q) \tag{225}$$

is precisely the index (9) of a 2d $\mathcal{N} = (0,2)$ vector multiplet for the gauge group $G = U(N)$. Let us note that the transformation between the continuous basis and the discrete basis is similar to the transformation considered in [GRRY11] where ordinary, non-affine characters were used.

If the Vafa–Witten partition function for the cobordism in the continuous basis is defined as

$$Z_{\text{VW}}^{U(N)}[B_{n+1,n}^{n+2,n+1}](q,y|z',z) = \sum_{\lambda,\rho} \chi_\lambda^{\hat{\mathfrak{u}}(N)_{n+2}}(q,z') \cdot Z_{\text{VW}}^{U(N)}[B_{n+1,n}^{n+2,n+1}]_\rho^\lambda(q,y) \cdot \chi_\rho^{\hat{\mathfrak{u}}(N)_{n+1}}(q,z) \cdot C_\rho^{-1}(q) \tag{226}$$

the relation (216) in the continuous basis should translate into the following property:

$$Z_{\text{VW}}^{U(N)}[A_{n+1}](q,\{y^{n+1},x_1/y,\dots,x_{n+1}/y\}|z') = \oint \prod_{j=1}^N \frac{dz_j}{2\pi i z_j}\, \mathcal{I}_V^{U(N)}(q,z)\, Z_{\text{VW}}^{U(N)}[B_{n+1,n}^{n+2,n+1}](q,y|z',z) \times Z_{\text{VW}}^{U(N)}[A_n](q,\{x_1,\dots,x_{n+1}\}|z) \tag{227}$$

or, explicitly,

$$\prod_{j=1}^N \left(\theta(-q^{\frac{1}{2}}y^{n+1}z_j';q)\prod_{i=1}^{n+1}\theta(-q^{\frac{1}{2}}x_i z_j'/y;q)\right) = \oint \prod_{j=1}^N \frac{dz_j}{2\pi i z_j}(q;q)_\infty^{2N}\prod_{i=1}^{n+1}\theta(-q^{\frac{1}{2}}x_i z_j;q)\prod_{i\neq j}\theta(z_i/z_j;q)\, Z_{\text{VW}}^{U(N)}[B_{n+1,n}^{n+2,n+1}](q,y|z',z). \tag{228}$$

The contour prescription is important and we take it to mean as evaluating the residue of the leading pole. If this is the case, then the following ansatz for $Z_{\text{VW}}^{U(N)}[B_{n+1,n}^{n+2,n+1}]$ solves Eq. (228):

$$Z_{\text{VW}}^{U(N)}[B_{n+1,n}^{n+2,n+1}](q,y|z',z) = \prod_{j=1}^N \theta(-q^{\frac{1}{2}}y^{n+1}z_j';q)\prod_{i,j=1}^N \frac{1}{\theta(z_i'/(z_j y);q)}. \tag{229}$$

The poles of the integral come from the denominator. They are at $z_i = z'_{\sigma(i)}/y$ for some permutation σ. After summing over all poles we end up with the desired result. From the form of the partition function we see that the cobordism corresponds to the theory of bifundamental chiral multiplets along with a fundamental Fermi multiplet. The Fermi multiplet itself can be associated with the 2-cycle in the cobordism which increases the second Betti number b_2 by 1.

Following the same reasoning one can deduce the partition function of the cobordism B transforming $A_{n_1-1} \sqcup \dots \sqcup A_{n_s-1} \rightsquigarrow A_{n_1+\dots+n_s-1}$. Consider $s = 2$

for simplicity. Then, $Z^{U(N)}_{\mathrm{VW}}[B]$ must satisfy

$$\begin{aligned} &Z^{U(N)}_{\mathrm{VW}}[A_{k+l-1}](q,\{y^l x_1,\ldots,x^l x_k, y^{-k}w_1,\ldots,y^{-k}w_l\}|z') \\ &\quad = \oint \prod_{j=1}^{N} \frac{dz_j}{2\pi i z_j}\frac{d\tilde{z}_j}{2\pi i \tilde{z}_j}\, \mathcal{I}^{U(N)}_V(q,z)\, \mathcal{I}^{U(N)}_V(q,\tilde{z})\, Z^{U(N)}_{\mathrm{VW}}[B](q,y|z',z,\tilde{z}) \\ &\qquad \times\ Z^{U(N)}_{\mathrm{VW}}[A_{k-1}](q,\{x_1,\ldots,x_k\}|z)\, Z^{U(N)}_{\mathrm{VW}}[A_{l-1}](q,\{w_1,\ldots,w_l\}|z) \end{aligned} \tag{230}$$

$$\begin{aligned} &\prod_{j=1}^{N}\prod_{i=1}^{k}\theta(-q^{\frac{1}{2}}y^l x_i z'_j;q)\prod_{i=1}^{l}\theta(-q^{\frac{1}{2}}x^{-k}w_i z'_j;q) \\ &\quad = \oint \prod_{j=1}^{N}\frac{dz_j}{2\pi i z_j}(q;q)^{2N}_{\infty}\prod_{i\neq j}\theta(z_i/z_j;q)\prod_{j=1}^{N}\prod_{i=1}^{k}\theta(-q^{\frac{1}{2}}x_i z_j;q) \\ &\qquad \times \oint \frac{d\tilde{z}_j}{2\pi i \tilde{z}_j}(q;q)^{2N}_{\infty}\prod_{i\neq j}\theta(\tilde{z}_i/\tilde{z}_j;q)\prod_{j=1}^{N}\prod_{i=1}^{l}\theta(-q^{\frac{1}{2}}w_i\tilde{z}_j;q) \\ &\qquad \times Z^{U(N)}_{\mathrm{VW}}[B](q,y|z',z,\tilde{z}) \end{aligned} \tag{231}$$

In this case, the following ansatz solves the equation:

$$Z^{U(N)}_{\mathrm{VW}}[B](q,y|z',z,\tilde{z}) = \prod_{i,j}\frac{1}{\theta(y^l z'_i/z_j;q)}\prod_{i,j}\frac{1}{\theta(y^{-k}z'_i/\tilde{z}_j;q)}. \tag{232}$$

As we can see, this is the index of two sets of bifundamental chiral multiplets, cf. [GGP13]. For a general cobordism $A_{n_1-1}\sqcup\ldots\sqcup A_{n_s-1} \rightsquigarrow A_{n_1+\ldots+n_s-1}$, the corresponding 2d $\mathcal{N}=(0,2)$ theory is that of s sets of bifundamental chiral multiplets.

3.11 Handle Slides

Another source of identities on the partition functions is handle slide moves described in Sect. 2. Consider the following simple example. First, let us note that since $L(p,p-1)\cong L(p,1)$ the cobordism B for

$$M_4(\overset{-p}{\bullet}) \quad \overset{B}{\rightsquigarrow} \quad M_4(\overset{-p}{\bullet}\!\!-\!\!\!-\!\!\overset{-1}{\bullet}) \tag{233}$$

is the same (although we glue along the different component of $\underset{\cdot}{B}$) as for

$$A_{p-2} \quad \overset{B}{\rightsquigarrow} \quad A_{p-1} \tag{234}$$

Therefore,

$$Z_{\mathrm{VW}}^{U(N)}[B]_{\rho}^{\lambda} \;=\; \chi_{\lambda,\rho}^{\widehat{\mathfrak{su}}(p)_N/\widehat{\mathfrak{su}}(p-1)_N}\,. \tag{235}$$

as we argued in Sect. 3.3. On the other hand, sliding a 2-handle gives the following relation, cf. (14):

$$M_4(\overset{-p}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-1}{\bullet}) \;\cong\; M_4(\overset{-(p-1)}{\bullet}\;\;\overset{-1}{\bullet})\,. \tag{236}$$

Taking into account (201) one can expect that

$$\sum_{\rho} \chi_{\lambda,\rho}^{\widehat{\mathfrak{su}}(p)_N/\widehat{\mathfrak{su}}(p-1)_N}\; Z_{\mathrm{VW}}^{U(N)}[M_4(\overset{-p}{\bullet})]_{\rho} \;=\; \tilde{\chi}^{\hat{\mathfrak{u}}(N)_1}\; Z_{\mathrm{VW}}^{U(N)}[M_4(\overset{-(p-1)}{\bullet})]_{\lambda}\,. \tag{237}$$

One can consider more complicated handle slides, for example:

$$\overset{-p}{\bullet}\!-\!\overset{-1}{\bullet} \quad\longrightarrow\quad \overset{-p}{\bullet}\!\!-\!\!\!-\!\!\overset{-(p-1)}{}\!\!-\!\!\!-\!\!\overset{-(p-1)}{\bullet} \quad\longrightarrow\quad \overset{-4p+3}{\bullet}\!\!-\!\!\!-\!\!\overset{-2(p-1)}{}\!\!-\!\!\!-\!\!\overset{-(p-1)}{\bullet} \tag{238}$$

which gives the equation

$$\sum_{\rho} Z_{\mathrm{VW}}^{U(N)}[B_{4p-3,1}^{p-1,1}]_{\rho}^{\lambda}\; Z_{\mathrm{VW}}^{U(N)}[M_4(\overset{-4p+3}{\bullet})]_{\rho} \;=\; \tilde{\chi}^{\hat{\mathfrak{u}}(N)_1}\; Z_{\mathrm{VW}}^{U(N)}[M_4(\overset{-(p-1)}{\bullet})]_{\lambda}\,.$$

4 Bottom-Up Approach: From 2d (0, 2) Theories to 4-Manifolds

As explained in Sect. 2, a 4-manifold M_4 with boundary $M_3 = \partial M_4$ defines a half-BPS (B-type) boundary condition in a 3d $\mathcal{N} = 2$ theory $T[M_3]$, such that the boundary degrees of freedom are described by a 2d $\mathcal{N} = (0,2)$ theory $T[M_4]$. Similarly, a cobordism between M_3^- and M_3^+ corresponds to a wall between 3d $\mathcal{N} = 2$ theories $T[M_3^-]$ and $T[M_3^+]$ or, equivalently (via the "folding trick"), to a B-type boundary condition in the theory $T[M_3^+] \times T[-M_3^-]$, etc.

Therefore, one natural way to approach the correspondence between 4-manifolds and 2d $(0,2)$ theories $T[M_4]$ is by studying half-BPS boundary conditions in 3d $\mathcal{N} = 2$ theories. For this, one needs to develop sufficient technology for constructing such boundary conditions, which will be the goal of the present section.

4.1 Chiral Multiplets and 3d Lift of the Warner Problem

The basic building blocks of 3d $\mathcal{N} = 2$ theories, at least those needed for building theories $T[M_3]$, are matter multiplets (chiral superfields) and gauge multiplets (vector superfields) with various interaction terms: superpotential terms, Fayet–Iliopoulos terms, Chern–Simons couplings, etc.

Therefore, we start by describing B-type boundary conditions in a theory of n chiral multiplets that parametrize a Kähler target manifold X. Examples of such boundary conditions were recently studied in [OY13] and will be a useful starting point for our analysis here. After reformulating these boundary conditions in a more geometric language, we generalize this analysis in a number of directions by including gauge fields and various interaction terms.

In order to describe boundary conditions that preserve $\mathcal{N} = (0, 2)$ supersymmetry on the boundary it is convenient to decompose 3d $\mathcal{N} = 2$ multiplets into multiplets of 2d $\mathcal{N} = (0, 2)$ supersymmetry algebra, see, e.g., [Witt93]. Thus, each 3d $\mathcal{N} = 2$ chiral multiplet decomposes into a bosonic 2d (0, 2) chiral multiplet Φ and a fermionic chiral multiplet Ψ, as illustrated in Table 4. Then, there are two obvious choices of boundary conditions that either impose Neumann conditions on Φ and Dirichlet conditions on Ψ, or vice versa. In the first case, the surviving (0, 2) multiplet parametrizes a certain holomorphic submanifold $Y \subset X$, whereas the second choice leads to left-moving fermions that furnish a holomorphic bundle $\mathcal{E}$ over Y. Put differently, a choice of a Kähler submanifold $Y \subset X$ determines a B-type boundary condition in a 3d $\mathcal{N} = 2$ sigma-model on X, such that 2d boundary theory is a (0, 2) sigma-model with the target space Y and a holomorphic bundle $\mathcal{E} = T_{X/Y}$, the normal bundle to Y in X:

$$\left.\begin{array}{l}\Phi_i \;:\; \text{Neumann}\\ \Psi_i \;:\; \text{Dirichlet}\end{array}\right\} \quad \Rightarrow \quad Y \subset X \tag{239}$$

$$\left.\begin{array}{l}\Phi_i \;:\; \text{Dirichlet}\\ \Psi_i \;:\; \text{Neumann}\end{array}\right\} \quad \Rightarrow \quad \mathcal{E} = T_{X/Y} \tag{240}$$

Now let us include superpotential interactions.

Table 4 Decomposition of $\mathcal{N} = (2, 2)$ superfields and couplings into (0, 2) superfields and couplings

$\mathcal{N} = (2, 2)$ supersymmetry	$\mathcal{N} = (0, 2)$ supersymmetry
Vector superfield (twisted chiral superfield)	Fermi + adjoint chiral (Λ, Σ)
Chiral superfield	Chiral + Fermi (Φ, Ψ)
Superpotential $\mathcal{W}(\Phi)$	(0, 2) superpotential $J = \frac{\partial \mathcal{W}}{\partial \Phi}$
Charge q_Φ	$E = i\sqrt{2}\, q_\Phi\, \Sigma\, \Phi$

4.2 3d Matrix Factorizations

In general, there are three types of holomorphic couplings in 2d $(0,2)$ theories that play the role of a superpotential. The first type already appears in the conditions that define bosonic and fermionic chiral multiplets:

$$\overline{D}_+\Phi_i = 0 \quad , \quad \overline{D}_+\Psi_j = \sqrt{2}E_j(\Phi) \tag{241}$$

Here, $E_j(\Phi)$ are holomorphic functions of chiral superfields Φ_i. The second type of holomorphic couplings $J^i(\Phi)$ can be introduced by the following terms in the action:

$$S_J \;=\; \int d^2x d\theta^+ \, \Psi_i J^i(\Phi) + c.c. \tag{242}$$

where, as in the familiar superpotential terms, the integral is over half of the superspace. In a purely two-dimensional $(0,2)$ theory, supersymmetry requires

$$\sum_i E_i J^i \;=\; 0 \tag{243}$$

However, if a 2d $(0,2)$ theory is realized on the boundary of a 3d $\mathcal{N}=2$ theory that has a superpotential $\mathcal{W}(\Phi)$, then the orthogonality condition $E \cdot J = 0$ is modified to

$$E(\Phi)\cdot J(\Phi) \;=\; \mathcal{W}(\Phi) \tag{244}$$

This modification comes from a three-dimensional analog of the "Warner problem" [War95], and reduces to it upon compactification on a circle. It also leads to a nice class of boundary conditions that are labeled by factorizations (or, "matrix factorizations") of the superpotential $\mathcal{W}(\Phi)$ and preserve $\mathcal{N}=(0,2)$ supersymmetry. For example, a 3d $\mathcal{N}=2$ theory with a single chiral superfield and a superpotential $\mathcal{W}=\phi^k$ has $k+1$ basic boundary conditions, with $(0,2)$ superpotential terms

$$J(\phi) = \phi^m \qquad , \qquad E(\phi) = \phi^{k-m} \qquad , \qquad m = 0, \ldots, k \tag{245}$$

To introduce the last type of holomorphic "superpotential" couplings in $(0,2)$ theories, we note that in 2d theories with $(2,2)$ supersymmetry there are two types of F-terms: the superpotential $\mathcal{W}$ and the twisted superpotential $\widetilde{\mathcal{W}}$. In a dimensional reduction from 3d, the latter comes from Chern–Simons couplings. The distinction between these two types of F-terms is absent in theories with only $(0,2)$ supersymmetry. In particular, they both correspond to couplings of the form (242) with $J = \frac{\partial \mathcal{W}}{\partial \Phi}$ or $\widetilde{J} = \frac{\partial \widetilde{\mathcal{W}}}{\partial \Sigma}$, except in the latter case one really deals with the field-dependent Fayet–Iliopoulos (FI) terms:

$$S_{FI} \;=\; \int d^2x d\theta^+ \, \Lambda_i \widetilde{J}^i(\Sigma, \Phi) + c.c. \tag{246}$$

where the Fermi multiplet Λ_i is the gauge field strength of the i-th vector superfield. The possibility of such holomorphic couplings is very natural from the (mirror) symmetry between the superpotential and twisted superpotential in $(2,2)$ models. However, the importance of such terms and, in particular, the fact that they can depend on *charged* chiral fields was emphasized only recently [MQSS12]. The novelty of these models is that classically they are not gauge invariant, but nevertheless can be saved by quantum effects. This brings us to our next topic.

4.3 Anomaly Inflow

Now we wish to explain that not only the coupling of a 2d $\mathcal{N} = (0,2)$ theory $T[M_4]$ to a 3d $\mathcal{N} = 2$ theory $T[M_3]$ on a half-space is convenient, but in many cases it is also necessary. In other words, by itself a 2d theory $T[M_4]$ associated with a 4-manifold with boundary may be anomalous. Such theories, however, do appear as building blocks in our story since the anomaly can be cancelled by inflow from the 3d space-time where $T[M_3]$ lives [CH85].

In this mechanism, the one-loop gauge anomaly generated by fermions in the 2d $(0,2)$ theory $T[M_4]$ is typically balanced against the boundary term picked up by anomalous gauge variation of the classical Chern–Simons action in 3d $\mathcal{N} = 2$ theory $T[M_3]$. Essentially the same anomaly cancellation mechanism—with Chern–Simons action in extra dimensions replaced by a WZW model—was used in a wide variety of hybrid $(0,2)$ models [GPS93, Joh95, BJKZ96, DS10, AG], where the chiral fermion anomaly and the classical anomaly of the gauged WZW model were set to cancel each other out. In particular, our combined 2d-3d system of theories $T[M_4]$ and $T[M_3]$ provides a natural home to the "fibered WZW models" of [DS10], where the holomorphic WZW component is now interpreted as Chern–Simons theory in extra dimension.

The simplest example—already considered in this context in [GGP13]—is an abelian 3d $\mathcal{N} = 2$ Chern–Simons theory at level k. In the presence of a boundary, it has k units of anomaly inflow which must be cancelled by coupling to an "anomalous heterotic theory"

$$\partial_\mu J^\mu \;=\; \frac{\mathcal{A}_R - \mathcal{A}_L}{2\pi} \alpha \epsilon^{\mu\nu} F_{\mu\nu} \tag{247}$$

whose left-moving and right-moving anomaly coefficients are out of balance by k units:

$$\mathcal{A}_R - \mathcal{A}_L \;=\; k \tag{248}$$

4.4 Boundary Conditions for $\mathcal{N} = 2$ Chern–Simons Theories

In general, there can be several contributions to the anomaly coefficients $\mathcal{A}_{L,R}$ and, correspondingly, different ways of meeting the anomaly cancellation condition like (248). In the case of a single $U(1)$ gauge symmetry, there is, of course, a familiar contribution from fermions transforming in chiral representations of the gauge group,

$$\mathcal{A}_R = \sum_{r:\text{chiral}} \tilde{q}_r^2 \tag{249a}$$

$$\mathcal{A}_L = \sum_{\ell:\text{Fermi}} q_\ell^2 \tag{249b}$$

where $\tilde{q}_r$ and q_ℓ are the charges of $(0,2)$ chiral and Fermi multiplets, respectively.

Besides the chiral anomaly generated by charged Weyl fermions, there can be an additional contribution to (248) from field-dependent Fayet–Iliopoulos couplings (246), such as "charged log interactions":

$$\widetilde{J} = \frac{i}{8\pi} \sum_r N_r \log(\Phi_r) \tag{250}$$

which spoils gauge invariance at the classical level. As explained in [MQSS12] such terms contribute to the anomaly

$$\Delta\mathcal{A}_R = -\sum_{r:\text{chiral}} \tilde{q}_r N_r \tag{251}$$

and arise from integrating out massive pairs of $(0,2)$ multiplets with unequal charges. Note the sign difference in (249a) compared to (251).

This can be easily generalized to a 2d-3d coupled system with gauge symmetry $U(1)^n$. Namely, let us suppose that 3d $\mathcal{N} = 2$ theory in this combined system contains Chern–Simons interactions with a matrix of "level" coefficients k_{ij}, much like our quiver Chern–Simons theory (37) associated with a plumbing graph Υ. And suppose that on a boundary of the 3d space-time it is coupled to some interacting system of $(0,2)$ chiral and Fermi multipets that, respectively, carry charges $\tilde{q}_r^i$ and q_ℓ^i under $U(1)^n$ symmetry, $i = 1, \ldots, n$. In addition, for the sake of generality we assume that the Lagrangian of the 2d $(0,2)$ boundary theory contains field-dependent FI terms (246) with

$$\widetilde{J}^i = \frac{i}{8\pi} \sum_r N_r^i \log(\Phi_r) \tag{252}$$

Then, the total anomaly cancellation condition for the coupled 2d-3d system—that combines all types of contributions (248), (249), and (251)—has the following form:

$$\sum_{r:\text{chiral}} \tilde{q}_r^i \tilde{q}_r^j - \sum_{\ell:\text{Fermi}} q_\ell^i q_\ell^j - \sum_{r:\text{chiral}} \tilde{q}_r^{(i} N_r^{j)} = k_{ij} \tag{253}$$

which must be satisfied for all values of $i, j = 1, \ldots, n$. Note that each of the contributions on the left-hand side can be viewed as a "matrix factorization" of the matrix of Chern–Simons coefficients. In particular, the term $\sum \tilde{q}_r^{(i} N_r^{j)}$ is simply the (symmetrized) product of the matrix of chiral multiplet charges and the matrix of the boundary superpotential coefficients, which altogether can be viewed as a "twisted superpotential version" of the condition (244), with (39) and (252).

Suppose for simplicity that we have a theory of free chiral and Fermi multiplets. The elliptic genus of this theory is simply

$$\mathcal{I}(q, x) = \frac{\prod_{\ell:\text{Fermi}} \theta(\prod_i x_i^{q_\ell^i}; q)}{\prod_{r:\text{chiral}} \theta(\prod_i x_i^{\tilde{q}_r^i}; q)} \tag{254}$$

In [BDP] it was argued that the right-hand side can be interpreted as the "half-index" of CS theory, that is, the partition function on $S^1 \times_q D$ which has boundary $S^1 \times_q S^1 \cong T^2$ with modulus τ. Following [GGP13] one can argue that this theory is equivalent to the quiver CS theory with coefficients k_{ij} living in the half-space on the left of 2d world-volume. That is, the original 2d-3d system is equivalent to CS theory in the whole space. The relation

$$k_{ij} = \sum_{r:\text{chiral}} \tilde{q}_r^i \tilde{q}_r^j - \sum_{\ell:\text{Fermi}} q_\ell^i q_\ell^j \tag{255}$$

can be deduced by considering the limit $q \to 1$ using that $\theta(x; q) \sim \exp\{-(\log x)^2/(2 \log q)\}$

Now, one can apply this to 3d $\mathcal{N} = 2$ theories $T[M_3; G]$ that come from fivebranes on 3-manifolds. Luckily, many of these theories—even the ones coming from multiple fivebranes, i.e., associated with non-abelian G—admit a purely abelian UV description, for which (253) should suffice. Hence, using the tools explained here one can match 4-manifolds to specific boundary conditions that preserve $\mathcal{N} = (0, 2)$ supersymmetry in two dimensions.

4.5 From Boundary Conditions to 4-Manifolds

Let us start with boundary conditions that can be described by free fermions. Clearly, these will give us the simplest examples of 2d $(0, 2)$ theories $T[M_4]$, some of which have been already anticipated from the discussion in the previous sections.

In particular, we expect to find free fermion description of theories $T[M_4(\Upsilon)]$ for certain plumbing graphs Υ. In the bottom-up approach of the present section, we construct such theories as boundary conditions in 3d $\mathcal{N} = 2$ theories $T[M_3]$ associated with $M_3 = \partial M_4$. Thus, aiming to produce a boundary condition for the $\mathcal{N} = 2$ quiver Chern–Simons theory (37), let us associate a symmetry group $U(1)_i$ to every vertex $i \in \Upsilon$ of the plumbing graph. Similarly, to every edge between vertices "i" and "j" we associate a Fermi multiplet carrying charges $(+1, -1)$ under $U(1)_i \times U(1)_j$. Then, its contribution to the gauge anomaly (253) is given by the matrix of anomaly coefficients that is non-trivial only in a 2×2 block (that corresponds to rows and columns with labels "i" and "j"):

$$-\mathcal{A}_L = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \tag{256}$$

To ensure cancellation of the total anomaly, a combination of such contributions must be set to equal the matrix of Chern–Simons coefficients k_{ij}, which for the quiver Chern–Simons theory (37) is given by the symmetric bilinear form (20). Therefore, by comparing (256) with (20), we immediately see that assigning $U(1)$ factors to vertices of the plumbing graph Υ and "bifundamental" charged Fermi multiplets to edges already accounts for all off-diagonal terms (with $i \neq j$) in the intersection form Q.

Also, note that contributions of charged Fermi multiplets to the diagonal elements of the anomaly matrix are always negative, no matter what combination of contributions (256) or more general charge assignments in (253) we take. This conclusion, of course, relies crucially on the signs in (253) and has an important consequence: only negative definite intersection forms Q can be realized by free Fermi multiplets.

For example, in the case of the A_n plumbing graph shown in Fig. 1, we have $M_3 = L(n+1, n)$, and the $\mathcal{N} = 2$ quiver Chern–Simons theory $T[L(n+1,n); U(1)]$ has matrix of Chern–Simons coefficients of the form (18) with $a_i = -2$, $i = 1, \ldots, n$. By combining (256) with two extra Fermi multiplets of charges ± 1 under the first and the last $U(1)$ factors, we can realize the A_n intersection form as the anomaly matrix in the following 2d $\mathcal{N} = (0, 2)$ theory:

$$T[M_4(A_n); U(1)] \;=\; \text{Fermi multiplets } \Psi_{\ell=0,\ldots,n} \tag{257}$$

with charges

$$q(\Psi_\ell) \;=\; \begin{cases} +1 \text{ under } U(1)_1, & \text{if } \ell = 0 \\ (-1, +1) \text{ under } U(1)_\ell \times U(1)_{\ell+1}, & \text{if } 1 \leq \ell < n \\ -1 \text{ under } U(1)_n, & \text{if } \ell = n \end{cases} \tag{258}$$

Note, the total number of Fermi multiplets in this theory is $n+1$, which is precisely the number of Taub-NUT centers in the ALE space of type A_n.

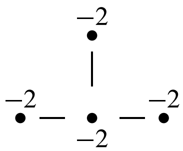
Let us briefly pause to discuss the structure of the charge matrix $(q^i_\ell)^{i=1,\dots,n}_{\ell=0,\dots,n}$ in (258). First, it is easy to see that each of the $U(1)^n$ gauge symmetries is "vector-like" in a sense that the charges add up to zero for every $U(1)$ factor. Also note that redefining the charges $q^n_\ell \mapsto q^1_\ell + 2q^2_\ell + 3q^3_\ell + \cdots + nq^n_\ell$ for all Fermi multiplets as in (102) gives a new matrix of charges that, via (253), leads to a new matrix of Chern–Simons coefficients:

$$Q \;=\; A_{n-1} \oplus \langle -n(n+1) \rangle \tag{259}$$

which splits into a matrix of Chern–Simons coefficients for a similar $U(1)^{n-1}$ theory and an extra $\mathcal{N} = 2$ Chern–Simons term at level $-n(n+1)$. In this basis we recognize the statement—explained in Sect. 2.14 through a variant of the "Norman trick" [Nor69, Qui79]—that a sphere plumbing with $\Upsilon = A_n$ can be built from the A_{n-1} sphere plumbing by a cobordism (attaching a 2-handle) with the intersection form $Q_B = \langle -n(n+1) \rangle$, cf. (103).

This observation has a nice physical interpretation in the coupled 2d-3d system described in Sect. 2.14 and illustrated in Figs. 14 and 15. Namely, the system of Fermi multiplets (257)–(258) without Ψ_n is simply the 2d $\mathcal{N} = (0,2)$ theory $T[M_4(A_{n-1}); U(1)]$ that can cancel anomaly and define a consistent boundary condition in the 3d $\mathcal{N} = 2$ Chern–Simons theory $T[M_3(A_{n-1}); U(1)]$ associated with the plumbing graph $\Upsilon = A_{n-1}$ by the general rule (37). In the new basis, the extra $U(1)_{i=n}$ symmetry (which is not gauged in $T[M_3(A_{n-1}); U(1)]$) is, in fact, an axial symmetry under which all $\Psi_{\ell=0,\dots,n-1}$ have charge $+1$. Gauging this symmetry and adding an extra Fermi multiplet that in the new basis has charge $-n$ under $U(1)_{i=n}$ gives precisely the 2d-3d system of 3d $\mathcal{N} = 2$ quiver Chern–Simons theory $T[M_3(A_n); U(1)]$ coupled to the 2d $\mathcal{N} = (0,2)$ theory $T[M_4(A_n); U(1)]$ on the boundary. This way of building $T[M_4(A_n); U(1)]$ corresponds to a fusion of the fully transmissive domain wall that carries Ψ_n with a boundary theory $T[M_4(A_{n-1}); U(1)]$, as illustrated in Figs. 14 and 15.

And, last but not least, in the matrix of charges $(q^i_\ell)^{i=1,\dots,n}_{\ell=0,\dots,n}$ given in (258) one can recognize simple roots $\alpha_{i=1,\dots,n}$ of the A_n root system. This suggests immediate generalizations. For instance, for a 4-manifold (105) whose plumbing graph $\Upsilon = D_4$ contains a trivalent vertex, we propose the "trinion theory" $T[\perp]$ to be a theory of four Fermi multiplets with the following charge assignments under the $U(1)^4$ flavor symmetry group:



$$: \qquad (q^i_\ell)_{\text{trinion}} \;=\; \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \tag{260}$$

The rows of this matrix are simple roots of the D_4 root system associated with the plumbing graph Υ, while the columns are the charge vectors of the Fermi multiplets $\Psi_{\ell=1,\dots,4}$. Substituting this into (253), we conclude that this 2d trinion theory can precisely cancel the anomaly of the 3d $\mathcal{N} = 2$ Chern–Simons theory with gauge group $U(1)^4$ and the matrix of Chern–Simons coefficients:

$$(Q_{ij}) = \begin{pmatrix} -2 & 1 & 0 & 0 \\ 1 & -2 & 1 & 1 \\ 0 & 1 & -2 & 0 \\ 0 & 1 & 0 & -2 \end{pmatrix} \tag{261}$$

which equals minus the Cartan matrix of the D_4 root system. This is in complete agreement with our general proposal (37) that $T[M_4(\Upsilon)]$ defines a consistent, non-anomalous boundary condition for the 3d $\mathcal{N} = 2$ theory $T[M_3(\Upsilon)]$, which in the present case is simply the quiver Chern–Simons theory defined by the symmetric bilinear form (20).

In Sect. 2.7 we saw that A_n linear plumbing can be naturally glued to a twisted D^2 bundle over S^2 with Euler number $-(n+1)$ since they share the same boundary (with opposite orientation, as required for gluing). In particular, the latter 4-manifold is represented by the Kirby diagram (15) with $p = n + 1$ and has boundary $M_3 = L(n+1, 1)$.

The corresponding 3d $\mathcal{N} = 2$ theory $T[L(n+1, 1); U(1)]$ was derived in (45): it is a $U(1)$ Chern–Simons theory at level $-(n+1)$. This theory can be related to the $U(1)^n$ quiver Chern–Simons theory $T[L(n+1, n); U(1)]$, cf. (51), by a sequence of dualities (3d Kirby moves) described in Sect. 2.7. In particular, this chain of dualities shows that $T[L(n+1, n); U(1)]$ and $T[L(n+1, 1); U(1)]$ are related by a parity transformation (65):

$$T[L(n+1, n)] \simeq P \circ T[L(n+1, 1)] \tag{262}$$

which, of course, is expected to hold for any G, not just $G = U(1)$.

Given the explicit description of the 3d $\mathcal{N} = 2$ theory $T[L(n+1, 1); U(1)]$, one can study B-type boundary conditions and try to match those with 4-manifolds bounded by $L(n+1, 1)$. The anomaly cancellation condition (253) suggests several possible candidates for the (0, 2) boundary theory $T[M_4]$:

(a) $n + 1$ Fermi multiplets of charge ± 1 (or, more generally, a collection of Fermi multiplets whose charges squared add up to $n + 1$);
(b) a single (0, 2) chiral multiplet Φ of charge $\tilde{q}_\Phi = +1$ and charged log interaction (252) with $N_\Phi = n + 2$.

4.6 Non-abelian Generalizations and Cobordisms

It is straightforward to extend this discussion to boundary theories and theories $T[M_4; G]$ trapped on walls for non-abelian G. Even if G is non-abelian, theories $T[M_4; G]$ and $T[M_3; G]$ often admit (multiple) UV definitions that only involve abelian gauge fields. In some cases, however, it is convenient to build $T[M_4; G]$ and $T[M_3; G]$ using non-abelian gauge symmetries. For instance, the Lens space theory (52) proposed in Sect. 2.2 is a good example.

In order to accommodate such examples, we need to discuss 2d $(0,2)$ theories with non-abelian gauge symmetries, which by itself is a very interesting subject that does not appear to be explored in the literature on $(0,2)$ heterotic models. Specifically, consider a general 2d theory with $(0,2)$ chiral multiplets Φ_r that transform in representations $\widetilde{R}_r$ of the gauge group G and Fermi multiplets Ψ_ℓ in representations R_ℓ. The corresponding fermions couple to the non-abelian gauge field via the usual covariant derivatives, e.g., for left-moving fermions in Fermi multiplets we have

$$(D_z)_{ij} \;=\; \delta_{ij}\partial_z + \sum_a A_z^a (T^a_{R_\ell})_{ij}$$

and similarly for chiral multiplets. Here, T^a_R are matrices of size $\dim(R)\times\dim(R)$ that obey the same commutation relations as the generators T^a of the Lie algebra Lie (G). (The latter correspond to the fundamental representation.) Then, the anomaly cancellation condition in such a theory has the form, cf. (253),

$$\sum_{r:\text{chiral}} \text{Tr}\,[T^a_{\widetilde{R}_r} T^b_{\widetilde{R}_r}] - \sum_{\ell:\text{Fermi}} \text{Tr}\,[T^a_{R_\ell} T^b_{R_\ell}] \;=\; (k^+ - k^-)\cdot \text{Tr}\,[T^a T^b] \tag{263}$$

where, in order to diversify our applications, we now assumed that the inflow from three dimensions has two contributions, from Chern–Simons couplings at levels k^+ and k^-, respectively. This more general form of the anomaly inflow is realized in a 2d $(0,2)$ theory trapped on a domain wall between 3d $\mathcal{N} = 2$ theories $T[M_3^+]$ and $T[M_3^-]$.

The anomaly cancellation condition (263) can be written more succinctly by using the index $C(R)$ of a representation R defined via Tr $\left(T^a_R T^b_R\right) = C(R)\delta^{ab}$. For example, for the fundamental and adjoint representations of $G = SU(N)$ we have $C(\text{fund}) = \frac{1}{2}$ and $C(\text{Adj}) = N$, respectively. In general,

$$C(R) \;=\; h_R \frac{\dim(R)}{\dim(\text{Adj})} \tag{264}$$

where h_R is the quadratic Casimir of the representation R.

Now we can apply (263), say, to the Lens space theory (52). We conclude that a domain wall that carries a Fermi multiplet Ψ in the fundamental representation of $G = U(N)$ changes the level of the $\mathcal{N} = 2$ Chern–Simons theory by one unit,

$$k^+ - k^- \;=\; -1 \tag{265}$$

This is consistent with our proposal, based on matching the Vafa–Witten partition function with the superconformal index, that the cobordism B that relates A_p and

A_{p+1} sphere plumbings corresponds to a domain wall which carries 2d $(0,2)$ theory

$$T[B;U(N)] \;=\; \text{Fermi multiplet } \Psi \text{ in the fundamental representation} \qquad (266)$$

The fusion of such domain walls is clearly non-singular and gives

$$T[M_4(A_p);U(N)] \;=\; p+1 \text{ Fermi multiplets } \Psi_{\ell=0,\dots,p} \text{ in } N\text{-dimn'l representation}$$

In fact, the wall in this example is fully transmissive. Notice, as in (257), the total number of Fermi multiplets in this theory is greater (by one) than the number of 2-handles in M_4 and equals the number of Taub-NUT centers in the ALE space of type A_p.

5 Future Directions

There are many avenues along which one can continue studying 2d $\mathcal{N} = (0,2)$ theories $T[M_4]$ labeled by 4-manifolds. The most obvious and/or interesting items on the list include:

- **Examples:** While focusing on the general structure, we presented a number of concrete (abelian and non-abelian) examples of: (a) theories labeled by 4-manifolds and 3-manifolds, (b) dualities that correspond to Kirby moves, (c) relations between cosets and Vafa–Witten partition functions, and (d) B-type walls and boundary conditions in 3d $\mathcal{N} = 2$ theories. Needless to say, it would certainly be interesting to extend our list of examples in each case.

  In particular, it would be interesting to study 2d $\mathcal{N} = (0,2)$ theories $T[M_4]$ associated with 4-manifolds that are not definite or not simply connected. Such examples clearly exist (e.g., for $M_4 = T^2 \times \Sigma_g$ or $M_4 = K3$, possibly with "frozen singularities" [Wit98, dDHKM02]), but still remain rather isolated and beg for a more systematic understanding, similar to theories labeled by a large class of negative definite simply connected 4-manifolds (13) considered in this paper. Thus, in Sect. 2 we briefly discussed a natural generalization to plumbings of twisted D^2 bundles over genus-g Riemann surfaces. It would be interesting to see what happens to the corresponding theories $T[M_4]$ when Riemann surfaces have boundaries/punctures and to make contact with [GRRY11].
- **4-manifolds with corners:** Closely related to the last remark is the study of 4-manifolds with corners. Although such situations were encountered at the intermediate stages in Sect. 2.2, we quickly tried to get rid of 3-manifolds with boundaries performing Dehn fillings. It would be interesting to study whether Vafa–Witten theory admits the structure of extended TQFT and, if it does, pursue the connection with gluing discussed in Sect. 2.2.
- **Smooth structures:** As was already pointed out in the introduction, it would be interesting to understand what the existence of a smooth structure on M_4 means for the corresponding 2d $\mathcal{N} = (0,2)$ theory $T[M_4]$. We plan to tackle this problem by studying surface operators in the fivebrane theory.

- **Large N limit:** It would be interesting to study the large N behavior of the Vafa–Witten partition function on plumbing 4-manifolds and make contact with holographic duals.
- **Non-abelian** $(0, 2)$ **models:** It appears that not much is known about non-abelian 2d $(0, 2)$ gauge dynamics. While in general abelian (gauge) symmetries suffice for building theories $T[M_4]$ and $T[M_3]$, in Sects. 2.2 and 4.5 we saw some examples where using non-abelian symmetries is convenient.
- **Defect junctions:** One important property of defect lines and walls is that they can form complicated networks and foam-like structures. Following the hints from Sects. 2.2–2.10 it would be interesting to understand if these play any role in the correspondence between 4-manifolds and 2d $(0, 2)$ theories.
- **Triangulations:** Since a basic d-dimensional simplex has $d + 1$ vertices, the Pachner moves in d dimensions involve adding one more vertex and then subdividing the resulting $(d+2)$-gon into basic simplices. In particular, for $d = 4$ such subdivisions always give a total of 6 simplices, resulting in 3–3 and 2–4 Pachner moves for 4-manifolds [Mac99]. It would be interesting to find a special function (analogous to the quantum dilogarithm for 2–3 Pachner moves in case of 3-manifolds) that enjoys such identities. Pursuing this approach, however, one should keep in mind that not every 4-manifold can be triangulated. Examples of non-triangulable 4-manifolds include some natural cases (such as Freedman's E_8 manifold mentioned in the Introduction) on which the fivebrane theory is expected to be well defined and interesting.

Acknowledgements We thank F. Quinn, D. Roggenkamp, C. Schweigert, A. Stipsicz, and P. Teichner for patient and extremely helpful explanations. We also thank T. Dimofte, Y. Eliashberg, A. Kapustin, T. Mrowka, W. Neumann, T. Okazaki, E. Sharpe, C. Vafa, J. Walcher, and E. Witten, among others, for a wide variety of helpful comments. The work of A.G. is supported in part by the John A. McCone fellowship and by DOE Grant DE-FG02-92-ER40701. The work of S.G. is supported in part by DOE Grant DE-FG03-92-ER40701FG-02 and in part by NSF Grant PHY-0757647. The work of P.P. is supported in part by the Sherman Fairchild scholarship and by NSF Grant PHY-1050729. Opinions and conclusions expressed here are those of the authors and do not necessarily reflect the views of funding agencies.

Appendix 1: M5-Branes on Calibrated Submanifolds and Topological Twists

We study the twisted compactification of 6d $(2, 0)$ theory on a four-manifold M_4. In each of the cases listed in Table 5, such compactification produces a superconformal theory $T[M_4]$ in the two non-compact dimensions. Via the computation of the T^2 partition function explained in the main text, the cases (a)–(c) correspond to previously studied topological twists of $\mathcal{N} = 4$ super-Yang-Mills which, in turn, are summarized in Table 6.

Specifically, in the first case (a) the $\mathcal{N} = 4$ SYM is thought of as an $\mathcal{N} = 2$ gauge theory with an extra adjoint multiplet and the Donaldson–Witten twist [Wit88]. Its

Table 5 Supersymmetric M5 brane compactifications on a negatively curved 4-manifold M_4

	R-symmetry $SO(5) \supset$	Embedding of M_4	SUSY	Solution	Metric on M_4
(a)	$SO(4) \supset SU(2) \times \boxed{SU(2)}$	Cayley in $Spin(7)$	$(0,1)$	$AdS_3 \times M_4$	Conf. half-flat
(b)	$\boxed{SO(4)}$	Lagrangian in CY_4	$(1,1)$	$AdS_3 \times M_4$	Const. curvature
(c)	$SO(2) \times \boxed{SO(3)}$	Coassociative in G_2	$(0,2)$	$AdS_3 \times M_4$	Conf. half-flat
(d)	$\boxed{SO(2)} \times \boxed{SO(2)}$	Kähler in CY_4	$(0,2)$	$AdS_3 \times M_4$	Kähler–Einstein
(e)	$SO(4) \supset U(2) \supset \boxed{U(1)}$	Kähler in CY_3	$(0,4)$	$AdS_3 \times S^2 \times CY_3$	Kähler–Einstein
(f)	$SO(4) \supset \boxed{U(2)}$	Complex Lagrangian in $d = 8$ hyper-Kähler	$(1,2)$	$AdS_3 \times M_4$	Kähler-Einstein w/ Const. hol. sec. curv.
(g)	$SO(4) \supset \boxed{SO(2)} \times \boxed{SO(2)}$	$(M_2 \subset CY_2) \times (M_2' \subset CY_2)$	$(2,2)$	$AdS_3 \times M_2 \times M_2'$	Const. curvature

In the first column we box the subgroup of $SO(5)$ R-symmetry of the M5 brane theory that is used to twist away the holonomy (or its subgroup) on M_4. Except in the case (e), all the AdS_3 solutions are already found in 7d supergravity and can be lifted to 11d by fibering S^4 over M_4, see, e.g., [GKW00, GK02, BB13]. In the case (e), the solution is found only in 11d supergravity. For manifolds M_4 with general holonomy (but still some restrictions on the metric), only the compactifications (a)–(c) are allowed. In this paper, we focus on the case (c) as it produces $(0,2)$ superconformal theory in two dimensions. In this case, M_4 is conformally half-flat; see, e.g., [Ito93] for moduli of conformally half-flat structures

Table 6 Topological twists of $\mathcal{N} = 4$ super-Yang-Mills

	R symmetry $SO(6) \supset$	Name	Equations
(a)	$SO(2) \times SU(2) \times \boxed{SU(2)}$	Donaldson–Witten	$F^+_{\alpha\beta} + [\overline{M}_{(\alpha}, M_{\beta)}] = 0$ $D_{\alpha\dot{\alpha}} M^\alpha = 0$
(b)	$SO(2) \times \boxed{SU(2)} \times \boxed{SU(2)}$	Marcus/GL	$F^+_{\mu\nu} - i[V_\mu, V_\nu]^+ = 0$ $(D_{[\mu} V_{\nu]})^- = 0 = D_\mu V^\mu$
(c)	$SO(3) \times \boxed{SO(3)}$	Vafa–Witten	$D_\mu C + \sqrt{2} D^\nu B^+_{\nu\mu} = 0$ $F^+_{\mu\nu} - \frac{i}{2}[B^+_{\mu\tau}, B^{+\tau}_{\nu}] - \frac{i}{\sqrt{2}}[B^+_{\mu\nu}, C] = 0$

path integral localizes on solutions to the non-abelian monopole equations. The untwisted rotation group of the DW theory is then twisted by the remaining $SU(2)$ symmetry to obtain the case (b). This twist (a.k.a. GL twist) was first considered by Marcus [Mar95] and related to the geometric Langlands program in [KW07]. The last case (c) is of most interest to us as it corresponds to $(0,2)$ SCFT in 2d. On a 4-manifold M_4, this twist is the standard Vafa–Witten twist [VW94].

Appendix 2: Orthogonality of Affine Characters

The Weyl–Kac formula for affine characters of $\widehat{\mathfrak{su}}(2)_k$ is

$$\chi_\lambda^{\widehat{\mathfrak{su}}(2)_k}(q,a) = \frac{\Theta^{(k+2)}_{\lambda+1}(a;q) - \Theta^{(k+2)}_{-\lambda-1}(a;q)}{\Theta^{(2)}_{1}(a;q) - \Theta^{(2)}_{-1}(a;q)} \tag{267}$$

where

$$\Theta^{(k)}_\lambda(a;q) := e^{-2\pi ikt} \sum_{n\in\mathbb{Z}+\lambda/2k} q^{kn^2} a^{kn} = e^{-2\pi ikt} q^{\frac{\lambda^2}{4k}} \sum_n q^{kn^2+\lambda n} a^{kn+\lambda} \tag{268}$$

Using the Weyl–Kac denominator formula the character can be rewritten as

$$\chi_\lambda^{\widehat{\mathfrak{su}}(2)_k}(q,a) = \frac{e^{-2\pi i(k+2)t} q^{\frac{(\lambda+1)^2}{4(k+2)}} \sum_n q^{(k+2)n^2} a^{(k+2)n} (q^{(\lambda+1)n} a^{(\lambda+1)} - q^{-(\lambda+1)n} a^{-(\lambda+1)})}{a^{-1}(q;q)\theta(a^2;q)}. \tag{269}$$

Consider the integral

$$\begin{aligned}
&\oint \frac{da}{2\pi ia} (q;q)^2_\infty \theta(a^2;q)\theta(a^{-2};q) \chi_\lambda^{\widehat{\mathfrak{su}}(2)_k}(q,a) \chi_{\lambda'}^{\widehat{\mathfrak{su}}(2)_k}(q,a) \\
&\quad = e^{-2\pi i(k+2)t} q^{\frac{(\lambda+1)^2}{4(k+2)} + \frac{(\lambda'+1)^2}{4(k+2)}} \\
&\qquad \times \sum_{n,m} \Bigg[q^{(k+2)(n^2+m^2)+(\lambda+1)n+(\lambda'+1)m} \oint \frac{da}{2\pi ia} a^{(k+2)(n+m)+(\lambda+1)+(\lambda'+1)} \\
&\qquad - q^{(k+2)(n^2+m^2)+(\lambda+1)n-(\lambda'+1)m} \oint \frac{da}{2\pi ia} a^{(k+2)(n-m)+(\lambda+1)-(\lambda'+1)} \\
&\qquad - q^{(k+2)(n^2+m^2)-(\lambda+1)n+(\lambda'+1)m} \oint \frac{da}{2\pi ia} a^{(k+2)(-n+m)-(\lambda+1)+(\lambda'+1)} \\
&\qquad + q^{(k+2)(n^2+m^2)-(\lambda+1)n-(\lambda'+1)m} \oint \frac{da}{2\pi ia} a^{(k+2)(-n-m)-(\lambda+1)-(\lambda'+1)} \propto \delta_{\lambda,\lambda'}
\end{aligned} \tag{270}$$

This shows that $\widehat{\mathfrak{su}}(2)_k$ characters are orthogonal with respect to the measure

$$(q;q)^2_\infty \theta(a^2;q)\theta(a^{-2};q) \tag{271}$$

but this measure is exactly the index of $SU(2)$ $(0,2)$ vector multiplet. The orthogonality of $\hat{\mathfrak{u}}(1)_k$ characters can be verified in a similar way. We conjecture that $\widehat{\mathfrak{su}}(N)_k$ ($\hat{\mathfrak{u}}(N)_k$) characters are orthogonal with respect to $SU(N)$ ($U(N)$) vector multiplet measure as well.

References

[Aus90] D.M. Austin, $SO(3)$-instantons on $L(p, q) \times \mathbf{R}$. J. Differ. Geom. **32**(2), 383–413 (1990)

[Ass96] T. Asselmeyer, Generation of source terms in general relativity by differential structures. Classical Quantum Gravity **14**, 749–758 (1997). [gr-qc/9610009]

[Akb12] S. Akbulut, 4-Manifolds. Oxford Graduate Texts in Mathematics, vol. 25 (Oxford University Press, Oxford, 2016)

[AG] A. Adams, D. Guarrera, Heterotic flux Vacua from hybrid linear models (2009) [arXiv:0902.4440]

[AG04] B.S. Acharya, S. Gukov, M theory and singularities of exceptional holonomy manifolds. Phys. Rep. **392**, 121–189 (2004). [hep-th/0409191]

[AV01] B.S. Acharya, C. Vafa, On domain walls of N=1 supersymmetric Yang-Mills in four-dimensions (2001). [hep-th/0103011]

[AW77] M.F. Atiyah, R.S. Ward, Instantons and algebraic geometry. Commun. Math. Phys. **55**(2), 117–124 (1977)

[ABT10] L.F. Alday, F. Benini, Y. Tachikawa, Liouville/Toda central charges from M5-branes. Phys. Rev. Lett. **105**, 141601 (2010). [arXiv:0909.4776]

[AGT10] L.F. Alday, D. Gaiotto, Y. Tachikawa, Liouville correlation functions from four-dimensional Gauge theories. Lett. Math. Phys. **91**, 167–197 (2010). [arXiv:0906.3219]

[APS73] M.F. Atiyah, V. Patodi, I. Singer, Spectral asymmetry and Riemannian geometry. Bull. Lond. Math. Soc **5**(2), 229–234 (1973)

[AOSV05] M. Aganagic, H. Ooguri, N. Saulina, C. Vafa, Black holes, q-deformed 2d Yang-Mills, and non-perturbative topological strings. Nucl. Phys. **B715**, 304–348 (2005). [hep-th/0411280]

[BB13] F. Benini, N. Bobev, Two-dimensional SCFTs from wrapped Branes and c-extremization. J. High Energy Phys. **1306**, 005 (2013). [arXiv:1302.4451]

[BM09] C. Bachas, S. Monnier, Defect loops in gauged Wess-Zumino-Witten models. J. High Energy Phys. **1002**, 003 (2010). [arXiv:0911.1562]

[BR07] I. Brunner, D. Roggenkamp, B-type defects in Landau-Ginzburg models. J. High Energy Phys. **0708**, 093 (2007). [arXiv:0707.0922]

[BT96] M. Blau, G. Thompson, Aspects of $N(T) \geq 2$ topological gauge theories and D-branes. Nucl. Phys. **B492**, 545–590 (1997). [hep-th/9612143]

[BT97] M. Blau, G. Thompson, Euclidean SYM theories by time reduction and special holonomy manifolds. Phys. Lett. **B415**, 242–252 (1997). [hep-th/9706225]

[BDP] C. Beem, T. Dimofte, S. Pasquetti, Holomorphic blocks in three dimensions. J. High Energy Phys. **2014**(12), Article 177, 118 pp. (2014)

[BJR08] I. Brunner, H. Jockers, D. Roggenkamp, Defects and D-Brane monodromies. Adv. Theor. Math. Phys. **13**, 1077–1135 (2009). [arXiv:0806.4734]

[BVS95] M. Bershadsky, C. Vafa, V. Sadov, D-branes and topological field theories. Nucl. Phys. **B463**, 420–434 (1996). [hep-th/9511222]

[BdDO02] C. Bachas, J. de Boer, R. Dijkgraaf, H. Ooguri, Permeable conformal walls and holography. J. High Energy Phys. **0206**, 027 (2002). [hep-th/0111210]

[BEHT13] F. Benini, R. Eager, K. Hori, Y. Tachikawa, Elliptic genera of two-dimensional N=2 gauge theories with rank-one gauge groups. Lett. Math. Phys. **104**(4), 465–493 (2014)

[BHKK99] O. Bergman, A. Hanany, A. Karch, B. Kol, Branes and supersymmetry breaking in three-dimensional gauge theories. J. High Energy Phys. **9910**, 036 (1999). [hep-th/9908075]

[BJKZ96] P. Berglund, C.V. Johnson, S. Kachru, P. Zaugg, Heterotic coset models and (0,2) string vacua. Nucl. Phys. **B460**, 252–298 (1996). [hep-th/9509170]

[CH85] C.G. Callan, J.A. Harvey, Anomalies and fermion zero modes on strings and domain walls. Nucl. Phys. **B250**, 427 (1985)

[CR10] N. Carqueville, I. Runkel, Rigidity and defect actions in Landau-Ginzburg models. Commun. Math. Phys. **310**, 135–179 (2012). [arXiv:1006.5609]

[CCV] S. Cecotti, C. Cordova, C. Vafa, Braids, walls, and mirrors (2011). [arXiv:1110.2115]

[Don83] S.K. Donaldson, An application of gauge theory to four-dimensional topology. J. Differ. Geom. **18**, 279–315 (1983)

[DS08] R. Dijkgraaf, P. Sulkowski, Instantons on ALE spaces and orbifold partitions. J. High Energy Phys. **0803**, 013 (2008). [arXiv:0712.1427]

[DS10] J. Distler, E. Sharpe, Heterotic compactifications with principal bundles for general groups and general levels. Adv. Theor. Math. Phys. **14**, 335–398 (2010). [hep-th/0701244]

[DGG1] T. Dimofte, D. Gaiotto, S. Gukov, Gauge theories labelled by three-manifolds. Commun. Math. Phys. **325**(2), 367–419 (2014)

[DGG2] T. Dimofte, D. Gaiotto, S. Gukov, 3-Manifolds and 3d Indices. [arXiv:1112.5179]

[DGG13] T. Dimofte, M. Gabella, A.B. Goncharov, K-Decompositions and 3d gauge theories (2013). [arXiv:1301.0192]

[DGH11] T. Dimofte, S. Gukov, L. Hollands, Vortex counting and Lagrangian 3-manifolds. Lett. Math. Phys. **98**, 225–287 (2011). [arXiv:1006.0977]

[DVV02] R. Dijkgraaf, E.P. Verlinde, M. Vonk, On the partition sum of the NS five-brane (2002). [hep-th/0205281]

[DHSV07] R. Dijkgraaf, L. Hollands, P. Sulkowski, C. Vafa, Supersymmetric gauge theories, intersecting Branes and free fermions. J. High Energy Phys. **0802**, 106 (2008). [arXiv:0709.4446]

[dDHKM02] J. de Boer, R. Dijkgraaf, K. Hori, A. Keurentjes, J. Morgan, et al., Triples, fluxes, and strings. Adv. Theor. Math. Phys. **4**, 995–1186 (2002). [hep-th/0103170]

[Fre82] M. Freedman, The topology of four dimensional manifolds. J. Differ. Geom. **17**, 357–453 (1982)

[FH90] M. Furuta, Y. Hashimoto, Invariant instantons on S^4. J. Fac. Sci. Univ. Tokyo Sect. IA Math. **37**(3), 585–600 (1990)

[FW99] D.S. Freed, E. Witten, Anomalies in string theory with D-branes. Asian J. Math. **3**, 819 (1999). [hep-th/9907189]

[FGP13] H. Fuji, S. Gukov, P. Sulkowski, Super-a-polynomial for knots and BPS states. Nucl. Phys. **B867**, 506–546 (2013). [arXiv:1205.1515]

[FSV12] J. Fuchs, C. Schweigert, A. Velentino, Bicategories for boundary conditions and for surface defects in 3-d TFT. Commun. Math. Phys. **321**(2), 543–575 (2013)

[FGSA] H. Fuji, S. Gukov, P. Sulkowski, H. Awata, Volume conjecture: refined and categorified. Adv. Theor. Math. Phys. **16**(2), 1669–1777 (2012)

[FGSS] H. Fuji, S. Gukov, M. Stošić, P. Sulkowski, 3d analogs of Argyres-Douglas theories and knot homologies. J. High Energy Phys. **2013**, 175 (2003)

[Gan96] O.J. Ganor, Compactification of tensionless string theories (1996). [hep-th/9607092]

[Guk05] S. Gukov, Three-dimensional quantum gravity, Chern-Simons theory, and the A-polynomial. Commun. Math. Phys. **255**(3), 577–627 (2005). [hep-th/0306165]

[Guk07] S. Gukov, Gauge theory and knot homologies. Fortschr. Phys. **55**, 473–490 (2007). [arXiv:0706.2369]

[Gai12] D. Gaiotto, N=2 dualities. J. High Energy Phys. **1208**, 034 (2012). [arXiv:0904.2715]

[GL91] T. Gannon, C. Lam, Gluing and shifting lattice constructions and rational equivalence. Rev. Math. Phys. **3**(03), 331–369 (1991)

[GL92] T. Gannon, C. Lam, Lattices and Θ-function identities. I: Theta constants. J. Math. Phys. **33**, 854 (1992)

[GL92] T. Gannon, C. Lam, Lattices and θ-function identities. II: Theta series. J. Math. Phys. **33**, 871 (1992)

[GK02] J.P. Gauntlett, N. Kim, M five-branes wrapped on supersymmetric cycles. 2.. Phys. Rev. **D65**, 086003 (2002). [hep-th/0109039]

[GK09] A. Giveon, D. Kutasov, Seiberg duality in Chern-Simons theory. Nucl. Phys. **B812**, 1–11 (2009). [arXiv:0808.0360]

[GP15] S. Gukov, D. Pei, Equivariant Verlinde formula from fivebranes and vortices (2015). [arXiv:1501.0131]

[GS99] R.E. Gompf, A.I. Stipsicz, 4-*manifolds and Kirby calculus*. Graduate Studies in Mathematics, vol. 20 (American Mathematical Society, Providence, RI, 1999)

[GW09] D. Gaiotto, E. Witten, Supersymmetric boundary conditions in N=4 super Yang-Mills theory. J. Stat. Phys. **135**, 789–855 (2009). [arXiv:0804.2902]

[GGP13] A. Gadde, S. Gukov, P.J. Putrov, Walls, lines, and spectral dualities in 3d Gauge theories. J. High Energy Phys. **2014**, 47 (2014)

[GKW00] J.P. Gauntlett, N. Kim, D. Waldram, M Five-branes wrapped on supersymmetric cycles. Phys. Rev. **D63**, 126001 (2001). [hep-th/0012195]

[GMN10] D. Gaiotto, G.W. Moore, A. Neitzke, Four-dimensional wall-crossing via three-dimensional field theory. Commun. Math. Phys. **299**, 163–224 (2010). [arXiv:0807.4723]

[GPS93] S.B. Giddings, J. Polchinski, A. Strominger, Four-dimensional black holes in string theory. Phys. Rev. **D48**, 5784–5797 (1993). [hep-th/9305083]

[GST02] S. Gukov, J. Sparks, D. Tong, Conifold transitions and five-brane condensation in M theory on spin(7) manifolds. Classical Quantum Gravity **20**, 665–706 (2003). [hep-th/0207244]

[GSW87] M.B. Green, J. Schwarz, E. Witten, *Superstring Theory. vol. 1: Introduction*, 1st edn. (Cambridge, New York, 1987)

[GVW00] S. Gukov, C. Vafa, E. Witten, CFT's from Calabi-Yau four folds. Nucl. Phys. **B584**, 69–108 (2000). [hep-th/9906070]

[GRRY11] A. Gadde, L. Rastelli, S.S. Razamat, W. Yan, The 4d superconformal index from q-deformed 2d Yang-Mills. Phys. Rev. Lett. **106**, 241602 (2011). [arXiv:1104.3850]

[Har79] J.L. Harer, *Pencils of Curves on 4-Manifolds* (ProQuest LLC, Ann Arbor, MI, 1979). Thesis (Ph.D.)-University of California, Berkeley

[HW97] A. Hanany, E. Witten, Type IIB superstrings, BPS monopoles, and three-dimensional gauge dynamics. Nucl. Phys. **B492**, 152–190 (1997). [hep-th/9611230]

[HW04] K. Hori, J. Walcher, D-branes from matrix factorizations. C. R. Phys. **5**, 1061–1070 (2004). [hep-th/0409204]

[Ito93] M. Itoh, Moduli of half conformally flat structures. Math. Ann. **296**(4), 687–708 (1993)

[Joh95] C.V. Johnson, Heterotic coset models. Mod. Phys. Lett. **A10**, 549–560 (1995). [hep-th/9409062]

[KP] V.G. Kac, D.H. Petersen, Infinite-dimensional Lie algebras, theta functions and modular forms. Adv. Math. **53**(2), 125–264 (1984)

[KS10] A. Kapustin, N. Saulina, Surface operators in 3d topological field theory and 2d rational conformal field theory, in *Mathematical Foundations of Quantum Field Theory and Perturbative String Theory.* Proceedings of Symposia in Pure Mathematics, vol. 83 (American Mathematical Society, Providence, 2011), pp. 175–198

[KS11] A. Kapustin, N. Saulina, Topological boundary conditions in abelian Chern-Simons theory. Nucl. Phys. **B845**, 393–435 (2011). [arXiv:1008.0654]

[KW07] A. Kapustin, E. Witten, Electric-magnetic duality and the geometric Langlands program. Commun. Num. Theor. Phys. **1**, 1–236 (2007). [hep-th/0604151]

[KW13] A. Kapustin, B. Willett, Wilson loops in supersymmetric Chern-Simons-matter theories and duality (2007). [arXiv:1302.2164]

[KOO99] T. Kitao, K. Ohta, N. Ohta, Three-dimensional gauge dynamics from brane configurations with (p,q) - five-brane. Nucl. Phys. **B539**, 79–106 (1999). [hep-th/9808111]

[Loc87] R. Lockhart, Fredholm, Hodge and Liouville theorems on noncompact manifolds. Trans. Am. Math. Soc. **301**(1), 1–35 (1987)

[LP72] F. Laudenbach, V. Poénaru, A note on 4-dimensional handlebodies. Bull. Soc. Math. Fr. **100**, 337–344 (1972)

[Mar95] N. Marcus, The other topological twisting of N=4 Yang-Mills. Nucl. Phys. **B452**, 331–345 (1995). [hep-th/9506002]

[Mac99] M. Mackaay, Spherical 2-categories and 4-manifold invariants. Adv. Math. **143**(2), 288–348 (1999)

[MNVW98] J. Minahan, D. Nemeschansky, C. Vafa, N. Warner, E strings and N=4 topological Yang-Mills theories. Nucl. Phys. **B527**, 581–623 (1998). [hep-th/9802168]

[MQSS12] I.V. Melnikov, C. Quigley, S. Sethi, M. Stern, Target spaces from Chiral gauge theories. J. High Energy Phys. **1302**, 111 (2013). [arXiv:1212.1212]

[Nor69] R.A. Norman, Dehn's lemma for certain 4-manifolds. Invent. Math. **7**, 143–147 (1969)

[Nak94] H. Nakajima, Instantons on ALE spaces, quiver varieties, and Kac-Moody Algebras. Duke Math. **76**, 365–416 (1994)

[NRXS12] S. Nawata, P. Ramadevi, Zodinmawia, X. Sun, Super-A-polynomials for twist knots. J. High Energy Phys. **1211**, 157 (2012). [arXiv:1209.1409]

[Oht99] K. Ohta, Supersymmetric index and s rule for type IIB branes. J. High Energy Phys. **9910**, 006 (1999). [hep-th/9908120]

[OA97] M. Oshikawa, I. Affleck, Boundary conformal field theory approach to the critical two-dimensional Ising model with a defect line. Nucl. Phys. **B495**, 533–582 (1997). [cond-mat/9612187]

[OY13] T. Okazaki, S. Yamaguchi, Supersymmetric boundary conditions in 3D N=2 theories, in *String-Math 2013*. Proceedings of Symposia in Pure Mathematics, vol. 88 (American Mathematical Society, Providence, 2014), pp. 343–349

[Pfe04] H. Pfeiffer, Quantum general relativity and the classification of smooth manifolds (2004). [gr-qc/0404088]

[Qui79] F. Quinn, Ends of maps. I. Ann. Math. (2) **110**(2), 275–331 (1979)

[Qui82] F. Quinn, Ends of maps. III. Dimensions 4 and 5. J. Differ. Geom. **17**(3), 503–521 (1982)

[QS02] T. Quella, V. Schomerus, Symmetry breaking boundary states and defect lines. J. High Energy Phys. **0206**, 028 (2002). [hep-th/0203161]

[Roh89] R. Rohm, Topological defects and Differential structures. Ann. Phys. **189**, 223 (1989)

[Sav02] N. Saveliev, Fukumoto-Furuta invariants of plumbed homology 3-spheres. Pac. J. Math. **205**(2), 465–490 (2002)

[Sla09] J. Sladkowski, Exotic smoothness and astrophysics. Acta Physiol. Pol. **B40**, 3157–3163 (2009). [arXiv:0910.2828]

[Smi10] A. Smilga, Witten index in supersymmetric 3d theories revisited. J. High Energy Phys. **1001**, 086 (2010). [arXiv:0910.0803]

[SW94] N. Seiberg, E. Witten, Electric - magnetic duality, monopole condensation, and confinement in N=2 supersymmetric Yang-Mills theory. Nucl. Phys. **B426**, 19–52 (1994). [hep-th/9407087]

[VW94] C. Vafa, E. Witten, A strong coupling test of S duality. Nucl. Phys. **B431**, 3–77 (1994). [hep-th/9408074]

[Wit87] E. Witten, Elliptic genera and quantum field theory. Commun. Math. Phys. **109**, 525 (1987)

[Wit88] E. Witten, Topological quantum field theory. Commun. Math. Phys. **117**, 353 (1988)

[Wit93] E. Witten, The Verlinde algebra and the cohomology of the Grassmannian (1993). [hep-th/9312104]
[Witt93] E. Witten, Phases of N=2 theories in two-dimensions. Nucl. Phys. **B403**, 159–222 (1993). [hep-th/9301042]
[Wit94] E. Witten, Monopoles and four manifolds. Math. Res. Lett. **1**, 769–796 (1994). [hep-th/9411102]
[War95] N. Warner, Supersymmetry in boundary integrable models. Nucl. Phys. **B450**, 663–694 (1995). [hep-th/9506064]
[Wit96] E. Witten, Five-brane effective action in M theory. J. Geom. Phys. **22**, 103–133 (1997). [hep-th/9610234]
[Wit98] E. Witten, Toroidal compactification without vector structure. J. High Energy Phys. **9802**, 006 (1998). [hep-th/9712028]
[Wit99] E. Witten, Supersymmetric index of three-dimensional gauge theory, in *The Many Faces of the Superworld* (World Scientific, River Edge, 2000), pp. 156–184
[Wit03] E. Witten, SL(2,Z) action on three-dimensional conformal field theories with Abelian symmetry, in *From Fields to Strings: Circumnavigating Theoretical Physics,* vol. 2 (World Scientific, Singapore, 2005), pp. 1173–1200
[WA94] E. Wong, I. Affleck, Tunneling in quantum wires: a boundary conformal field theory approach. Nucl. Phys. **B417**, 403–438 (1994)

Higgs Bundles and Characteristic Classes

Nigel Hitchin

Dedicated to the Memory of Friedrich Hirzebruch

1 Introduction

Sixty years ago Hirzebruch observed how the vanishing of the Stiefel–Whitney class w_2 led to integrality of the $\hat{A}$-genus of an algebraic variety [Hirz1]. This was one motivation for the Atiyah–Singer index theorem but also for my own thesis about Dirac operators and Kähler manifolds. Indeed the interaction between topology and algebraic geometry which he developed has been a constant theme in virtually all my work.

This article is also about w_2, characteristic classes and algebraic geometry, but in a rather different context. We consider a compact oriented surface Σ of genus g, a real Lie group G^r and the character variety $\mathrm{Hom}(\pi_1(\Sigma), G^r)/G^r$, equivalently the moduli space of flat G^r-connections on Σ. If U is the maximal compact subgroup of G^r, then each principal G^r-bundle has a characteristic class in $H^2(\Sigma, \pi_1(U))$ which helps to determine in which connected component of the character variety it lies.

A well-known example is the case $G^r = SL(2, \mathbf{R})$ where we have $U = SO(2)$ and a class $c \in H^2(\Sigma, \pi_1(SO(2))) \cong \mathbf{Z}$. It satisfies the Milnor–Wood inequality $|c| \leq 2g - 2$ and there is one component for each c for which strict inequality holds, but when $|c| = 2g - 2$ there are 2^{2g} connected components, each a copy of Teichmüller space.

Here we shall consider the groups $SL(n, \mathbf{R})$ and $Sp(2m, \mathbf{R})$, higher dimensional generalizations of $SL(2, \mathbf{R}) = Sp(2, \mathbf{R})$. In the first case we have a characteristic class in $H^2(\Sigma, \pi_1(SO(n)))$ which, for $n > 2$, is a Stiefel–Whitney class $w_2 \in \mathbf{Z}_2$ and in the second a Chern class c_1 in $H^2(\Sigma, \pi_1(U(m))) \cong \mathbf{Z}$.

We approach this question using the moduli space of Higgs bundles. For a complex group G^c and a complex structure on Σ, gauge-theoretic equations enable us to describe the G^c-character variety in terms of a holomorphic principal bundle

N. Hitchin (✉)
Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road,
Oxford, OX2 6GG, UK
e-mail: hitchin@maths.ox.ac.uk

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_8

and a holomorphic section Φ of $\mathfrak{g} \otimes K$ where K is the canonical bundle. This moduli space has the structure of a completely integrable Hamiltonian system—a proper map to an affine space, whose generic fibre is an abelian variety. The G^r-character variety is realized as the fixed point set of a holomorphic involution and for the real groups in question the involution acts trivially on the base and its fixed points can be identified with the elements of order 2 in the abelian variety. The main problem we address is to evaluate the characteristic class as a function on this $\mathbf{Z}_2$-vector space. For $SL(n, \mathbf{R})$, w_2 is a quadratic function related to the mod 2 index theorem and for $Sp(2m, \mathbf{R})$ the characteristic class is determined by the orbit of an action of a symmetric group under its permutation representation over $\mathbf{Z}_2$.

The above results provide the background for testing the predictions of mirror symmetry for the hyperkähler Higgs bundle moduli space, and this we approach in the final section of this paper. A component of the G^r-character variety is an example of what is known as a BAA-brane. The SYZ approach to mirror symmetry says that its mirror should be a BBB-brane. In this context, a BBB-brane is a hyperholomorphic bundle over a hyperkähler submanifold. Now it is known that mirror symmetry for these moduli spaces is closely related to Langlands duality and the duality of the abelian varieties in the integrable system. How this works for the two real forms above is still a mystery, but we shall indicate a conjectural mirror for the real form $G^r = U(m, m) \subset GL(2m, \mathbf{C})$. There is again an integral characteristic class here and our candidate for the mirror is, for each allowable value, a hyperholomorphic bundle over the moduli space of $Sp(2m, \mathbf{C})$-Higgs bundles considered as a hyperkähler submanifold of the $GL(2m, \mathbf{C})$-moduli space. It is trivial only for the components of the moduli space where the characteristic class takes its maximum absolute value.

2 Higgs Bundles

We summarize here the basic facts about Higgs bundles [Hit1, Sim]. A crucial feature is the hyperkähler structure which provides the non-holomorphic isomorphism between the character variety and the moduli space of Higgs bundles. This arises from an infinite-dimensional quotient construction.

Let Σ be a compact oriented Riemann surface of genus $g > 1$ and P a principal bundle for the compact real form G of a complex semi-simple Lie group G^c. The space of connections on P is an affine space $\mathcal{A}$ with group of translations $\Omega^1(\Sigma, \mathfrak{g})$ and a symplectic form given by integrating $B(a \wedge b)$ where B is an invariant metric on G. The complex structure on Σ gives this the structure of an infinite-dimensional flat Kähler manifold with complex tangent space $\Omega^{0,1}(\Sigma, \mathfrak{g})$. The group $\mathcal{G}$ of gauge transformations acts isometrically. The cotangent bundle $T^*\mathcal{A} = \mathcal{A} \times \Omega^{1,0}(\Sigma, \mathfrak{g})$ is a flat *hyperkähler* manifold and the induced gauge group action has a hyperkähler moment map, which applied to $(A, \Phi) \in \mathcal{A} \times \Omega^{1,0}(\Sigma, \mathfrak{g})$ is

$$\mu(A, \Phi) = (F_A + [\Phi, \Phi^*], \bar{\partial}_A \Phi), \tag{1}$$

where F_A is the curvature and $\Phi \mapsto \Phi^*$ for a general group is the antiholomorphic involution coming from the compact real form.

The zero set gives firstly $\bar{\partial}_A \Phi = 0$, so the Higgs field Φ is a holomorphic section of $\mathfrak{g} \otimes K$, and secondly the equation $F_A + [\Phi, \Phi^*] = 0$ which is equivalent to a stability condition. The quotient of this zero set by $\mathcal{G}$ is the moduli space of pairs (A, Φ), and it has an induced hyperkähler structure. This is a metric compatible with complex structures I, J, K satisfying the relations of the quaternions. If I denotes the complex structure of pairs (A, Φ), then J, K are the complex structures for the G^c-connections $\nabla_A + \Phi + \Phi^*$, $\nabla_A + i\Phi - i\Phi^*$, respectively. Setting (1) to zero shows that these are flat connections and by considering the holonomy, the moduli space with complex structure J or K can be identified with the G^c-character variety.

The integrable system for $G^c = GL(n, \mathbf{C})$ is defined by the characteristic polynomial of the Higgs field in its defining representation: $\det(x - \Phi) = x^n + a_1 x^{n-1} + \cdots + a_n$ where $a_i \in H^0(\Sigma, K^i)$. This maps the moduli space $\mathcal{M}$ to the vector space $\bigoplus_{1 \leq i \leq n} H^0(\Sigma, K^i)$: it is proper and the functions defined by it Poisson-commute with respect to the natural symplectic structure. The generic fibre is then a complex torus but it can be identified with the Jacobian of the curve with equation $\det(x - \Phi) = 0$. For complex linear groups the fibres correspond to certain abelian subvarieties of the Jacobian.

To obtain the character variety for the real form G^r in the Higgs bundle realization, we need the connection A to have holonomy in U, the maximal compact subgroup of G^r and, with $\mathfrak{g} = \mathfrak{u} \oplus \mathfrak{m}$, the Higgs field must lie in $H^0(\Sigma, \mathfrak{m} \otimes K)$. So for $G^r = SL(n, \mathbf{R})$, we have $U = SO(n)$ and the Higgs bundle is defined by a rank n holomorphic vector bundle V with an orthogonal structure and $\Lambda^n V$ trivial. The Higgs field Φ must then be symmetric with respect to this inner product [Hit3]. The characteristic class here is $w_2 \in \mathbf{Z}_2$: the obstruction to lifting the $SO(n, \mathbf{C})$-frame bundle to $\mathrm{Spin}(n, \mathbf{C})$.

For $G^r = Sp(2m, \mathbf{R})$, the maximal compact is $U = U(m)$ and this means the vector bundle $V = W \oplus W^*$ with W a rank m vector bundle [PG, GGM], the pairing between W and W^* defining the symplectic structure. Here the Higgs field has the off-diagonal form $\Phi(w, \xi) = (\beta(\xi), \gamma(w))$ where $\beta : W^* \to W \otimes K$ and $\gamma : W \to W^* \otimes K$ are symmetric. The characteristic class here is $c_1(W) \in \mathbf{Z}$.

3 The Canonical Section

As shown in [Hit3], there are canonical sections of the integrable system, each point of which gives a Higgs bundle for the split real form G^r. In particular, this gives a distinguished point in the generic fibre which we can regard as the identity element in an abelian variety. We spell this out next in our two cases which are indeed split real forms.

For $SL(n, \mathbf{R})$ and $n = 2m + 1$ the vector bundle is given by

$$V = K^{-m} \oplus K^{1-m} \oplus \cdots \oplus K^m$$

and for $n = 2m$ by

$$V = K^{-(2m-1)/2} \oplus K^{1-(2m-1)/2} \oplus \cdots \oplus K^{(2m-1)/2},$$

where for n odd we have to choose a square root of the canonical bundle K (a *theta characteristic* in classical terms, or a spin structure [MFA1] in the language of topology). The pairing of $K^{\pm\ell}$ or $K^{\pm\ell/2}$ defines an orthogonal structure on V and $\Lambda^n V$ is trivial so it has structure group $SO(n)$.

The subbundle $K^{1/2} \oplus \cdots \oplus K^{(2m-1)/2}$ when $n = 2m$ or $K \oplus \cdots \oplus K^m$ when $n = 2m+1$ is maximal isotropic and a spin structure for V is defined by a holomorphic square root of the top exterior power of a maximal isotropic subbundle. This in the two cases is $K^{m^2/2}$ and $K^{m(m+1)/2}$. These have Chern classes $m^2(g-1)$ and $m(m+1)(g-1)$. The latter is even and so in odd dimensions $w_2 = c_1 \bmod 2 = 0$. When $n = 2m$, $w_2 = 0$ if g is odd and if g is even $w_2 = m \bmod 2$.

The Higgs field must be symmetric with respect to this orthogonal structure. We set:

$$\Phi = \begin{pmatrix} 0 & 1 & 0 & \dots & & 0 \\ a_2 & 0 & 1 & \dots & & 0 \\ a_3 & a_2 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & & \vdots \\ a_{n-1} & & & & \ddots & 1 \\ a_n & a_{n-1} & \dots & a_3 & a_2 & 0 \end{pmatrix}, \tag{2}$$

where $a_i \in H^0(\Sigma, K^i)$.

Remark. In [Hit3, p. 456] it was claimed that this is conjugate to the companion matrix of the polynomial $x^n + a_2 x^{n-2} + \cdots + a_n$ which is incorrect. However, the coefficients of the characteristic polynomial are universal polynomials in the a_i and can be thought of as simply changing the basis of invariant polynomials on $\mathfrak{sl}(n)$. The actual characteristic polynomial can be viewed as follows (see [Tr]). Set $p(x) = 1 - \lambda x + a_2 x^2 + \cdots + a_n x^n$, then $p(x)$ and x^n have no common factor so there are unique polynomials $a(x), b(x)$ of degree $\leq (n-1)$ such that $a(x)p(x) + b(x)x^n = 1$. Then $b(0) = \det(\lambda - \Phi)$.

For $Sp(2m, \mathbf{R})$ the vector bundle is given by

$$V = K^{-(2m-1)/2} \oplus K^{1-(2m-1)/2} \oplus \cdots \oplus K^{(2m-1)/2}$$

where now we use the pairing between $K^{\pm\ell/2}$ to define a symplectic structure. Putting $W = K^{(2m-1)/2} \oplus K^{(2m-1)/2-2} \oplus \cdots \oplus K^{-(2m-3)/2}$ gives the form $V = W \oplus W^*$ above. Then $c_1(W) = m(g-1)$.

We need for the Higgs field sections β, γ of $S^2 W \otimes K, S^2 W^* \otimes K$, respectively. Since $W = W^* \otimes K$, we set $\gamma = 1$. Then we take for Φ the matrix of the form

$$\Phi = \begin{pmatrix} 0 & 1 \\ A & 0 \end{pmatrix},$$

where

$$A = \begin{pmatrix} a_2 & 1 & 0 & \dots & & 0 \\ a_4 & a_2 & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & & \vdots \\ & & & & \ddots & \\ & & & & & 1 \\ a_{2m} & & \dots & a_6 & a_4 & a_2 \end{pmatrix}. \tag{3}$$

Then $\det(\lambda - \Phi) = \det(\lambda^2 - A)$ and the coefficients are again universal polynomials in the a_i, providing another basis for the invariant polynomials. In fact, since $Sp(2m, \mathbf{R}) \subset SL(2m, \mathbf{R})$ the automorphisms of V generated by the a_i take the Higgs field for $Sp(2m, \mathbf{R})$ above into one of the form (2).

4 Spectral Data

Given a Higgs bundle for a linear group the spectral curve S is defined by the characteristic equation $0 = \det(x - \Phi) = x^n + a_1 x^{n-1} + \dots + a_n$, where $a_i \in H^0(\Sigma, K^i)$. It is a divisor of the line bundle $\pi^* K^n$ on the total space of K, where x is the tautological section of $\pi^* K$ on K, and since the canonical bundle of a cotangent bundle is trivial, $K_S \cong \pi^* K^n$ by adjunction. In particular the genus of S is given by $g_S - 1 = n^2(g-1)$. When S is smooth, the cokernel of $\pi^*\Phi - xI$ on S defines a line bundle $L\pi^*K (= L \otimes \pi^* K)$ and the vector bundle V can be recovered as $V = \pi_* L$, the direct image sheaf. The direct image of $x : L \to L\pi^* K$ is then the Higgs field Φ. The direct image of the trivial bundle is $\mathcal{O} \oplus K^{-1} \oplus K^{-2} \oplus \dots \oplus K^{-(n-1)}$ [BNR].

In general $\Lambda^n V \cong \mathrm{Nm}(L) K^{-n(n-1)/2}$ [BNR] where $\mathrm{Nm} : \mathrm{Pic}(S) \to \mathrm{Pic}(\Sigma)$ is the norm map which associates to a divisor $\sum n_i p_i$ on S the divisor $\sum n_i \pi(p_i)$ on Σ. Thus, to get an $SL(n, \mathbf{C})$-Higgs bundle we take $L \cong U \pi^* K^{(n-1)/2}$ where U lies in the Prym variety, the kernel of the homomorphism $\mathrm{Nm} : \mathrm{Pic}^0(S) \to \mathrm{Pic}^0(\Sigma)$. The canonical section described above is obtained by taking U to be the trivial bundle.

For the symplectic group $Sp(2m, \mathbf{C})$, the eigenvalues of Φ occur in pairs $\pm\lambda$ and the equation for the spectral curve has the form $x^{2m} + a_2 x^{2m-2} + \dots + a_{2m} = 0$. Thus S has an involution $\sigma(x) = -x$. In this case the bundle U must satisfy $\sigma^* U \cong U^*$ [Hit2]. This is the Prym variety for the map to the quotient $S \to \bar{S} = S/\sigma$.

These are the spectral data for the complex groups, next we need to find the restrictions for the real forms. For the group $SL(m, \mathbf{R})$, we need V to be orthogonal and Φ to be symmetric. This is a fixed point of a holomorphic involution on the Higgs bundle moduli space: $(V, \Phi) \mapsto (V^*, \Phi^T)$. Since the real dimension of $(\mathrm{Hom}(\pi_1(\Sigma), SL(m,\mathbf{R}))/SL(m,\mathbf{R})$ is $2(g-1)\dim SL(m,\mathbf{R})$, each component of the fixed point set has complex dimension $\dim \mathcal{M}/2$. The canonical Higgs bundle lies in the fixed point set as we have seen, and so the $+1$-eigenspace of the action on the tangent space at this point has dimension $\dim \mathcal{M}/2$. However, $\det(x-\Phi^T) = \det(x-\Phi)$ so the involution acts trivially on the base of the integrable system, which means the action on the tangent space to the fibre is -1. Since the fibre is known to be connected, by exponentiation the fixed points correspond to the elements of order 2 in the Prym variety. In fact, this argument holds for any split real form and is dealt with in [LS].

To see more concretely how the direct image $\pi_* L$ acquires an orthogonal structure when U^2 is trivial we use relative duality, with the equivalent condition $L^2 \cong \pi^* K^{n-1} \cong K_S \pi^* K^*$.

Relative duality in our situation states that for any vector bundle W on S, $(\pi_* W)^* \cong \pi_*(W^* \otimes K_S \pi^* K^*)$. Explicitly, over a regular value p of π,

$$(\pi_* W)_p = \bigoplus_{\pi(u)=p} W_u$$

and at each point $u \in S$ we have the derivative $d\pi_u \in (K_S \pi^* K^*)_u$. Then given $v \in (\pi_* W)_p$, $\xi \in \pi_*(W^* \otimes K_S \pi^* K^*)_p$ the non-degenerate pairing is

$$\langle v, \xi \rangle = \bigoplus_{\pi(u)=p} \frac{\xi(v)_u}{d\pi_u}.$$

At a branch point with the local form $z \mapsto w = z^k$ we write a local holomorphic section of $\pi_* W$ as $f(z) = b_0(w) + z b_1(w) + \cdots + z^{k-1} b_{k-1}(w)$ and then, if $g(z) = c_0(w) + z c_1(w) + \cdots + z^{k-1} c_{k-1}(w)$ is a local section of $\pi_*(W^* \otimes K_S \pi^* K^*)$ we have a contribution of

$$\lim_{z \to 0} \sum_{i=0}^{k-1} \frac{1}{k \omega^{-i} z^{k-1}} \langle f(\omega^i z), g(\omega^i z) \rangle = \sum_{j+\ell=k-1} \langle b_j, c_\ell \rangle,$$

where ω is a primitive kth root of unity.

So, returning to the case $L^2 \cong K_S \pi^* K^*$, the duality $V \cong V^*$ is expressed by the quadratic form

$$(s,s)_p = \bigoplus_{\pi(u)=p} \frac{s_u^2}{d\pi_u} \tag{4}$$

which is naturally a sum of squares over regular values. The Higgs field is the direct image of $s \mapsto xs$ and since $(xs)t_u = s(xt)_u$, Φ is symmetric.

In the symplectic case $\sigma^* U \cong U^*$ and since $L \cong U \pi^* K^{(n-1)/2}$ we have

$$\sigma^* L \cong L^* \pi^* K^{n-1} \cong L^* K_S \pi^* K^*.$$

Here, we have a non-degenerate bilinear form

$$\langle v, w\rangle = \bigoplus_{\pi(u)=p} \frac{\sigma^* v(w)}{d\pi_u} \tag{5}$$

which is skew-symmetric since $d\pi$ has opposite signs at $u, \sigma(u)$.

We also have the condition that U^2 is trivial and so $L^2 \cong K_S \pi^* K^*$. This means that $\sigma^* L \cong L$ and we have an action of σ (well-defined modulo ± 1) on L. Given an open set $A \subset \Sigma$, $\pi^{-1}(A) \subset S$ is invariant under σ and this means that the decomposition of $H^0(\pi^{-1}(A), L)$ into invariant and anti-invariant parts descends to a decomposition $V = W_1 \oplus W_2$. Since σ interchanges in pairs the $2m$ fibres, $\mathrm{rk}\, W_1 = \mathrm{rk}\, W_2 = m$.

Now if $s, t \in W_1$ they are represented by local invariant sections of L. Then from (5), $(\sigma^* s)t$ is invariant but the denominator $d\pi_u$ is anti-invariant and hence W_1 is Lagrangian, and hence $W_2 \cong W_1^*$. We therefore have the required form for $V = W \oplus W^*$. Now since $\sigma(x) = -x$, Φ interchanges W and W^* and as before $(xs)t = s(xt)$ is symmetric.

5 Characteristic Classes for $SL(n, \mathbf{R})$

In the previous section we saw how the direct image of a line bundle U of order 2 on the curve S defines an orthogonal bundle V on Σ. There are two characteristic classes $w_1(V)$ and $w_2(V)$ but $w_1 = 0$ if U lies in the Prym variety of $\pi : S \to \Sigma$. Topologically this means that if we take the dual homology class $u \in H_1(S, \mathbf{Z}_2)$ of $U \in H^1(S, \mathbf{Z}_2)$, then $\pi_*(u) = 0$.

The second Stiefel–Whitney class is more complicated.

To discuss the topology of orthogonal bundles on a surface Σ we use *KO*-theory, following [MFA1]. For a compact surface Σ

$$KO(\Sigma) \cong \mathbf{Z} \oplus H^1(\Sigma, \mathbf{Z}_2) \oplus H^2(\Sigma, \mathbf{Z}_2),$$

where the total Stiefel–Whitney class $w = 1 + w_1 + w_2$ gives an isomorphism of the additive group $\tilde{KO}(\Sigma)$ to the multiplicative group $1 \oplus H^1(\Sigma, \mathbf{Z}_2) \oplus H^2(\Sigma, \mathbf{Z}_2)$.

Generators are given by holomorphic line bundles L such that $L^2 \cong \mathcal{O}$ and the class $\Omega = \mathcal{O}_p + \mathcal{O}_p^* - 2$ where $\mathcal{O}_p$ is the holomorphic line bundle given by a point $p \in \Sigma$. We write $\alpha(x) \in KO(\Sigma)$ for the class of the line bundle corresponding to $x \in H^1(\Sigma, \mathbf{Z}_2)$. Then $\alpha(0) = 1$ and

$$\alpha(x + y) = \alpha(x) + \alpha(y) - 1 + (x, y)\Omega$$

This is nonlinear as it corresponds to the tensor product of line bundles. We have $w_1(\alpha(x)) = x, w_1(\Omega) = 0, w_2(\Omega) = c_1(\mathcal{O}_p) \bmod 2 = [\Sigma]$.

For an arbitrary rank n orthogonal bundle V we have

$$[V] = n - 1 + \alpha(w_1(V)) + w_2(V)\Omega\,.$$

Now choose a theta characteristic $K^{1/2}$ on Σ. This is a spin structure and defines a KO-orientation. The map to a point gives an invariant which is a spin cobordism characteristic number, an additive homomorphism $\varphi : KO(\Sigma) \to \mathbf{Z}_2$.

Given a holomorphic bundle V with an orthogonal structure, φ is an analytic mod 2 index $\varphi([V]) = \dim H^0(\Sigma, V \otimes K^{1/2})$ mod 2. For $V = \mathcal{O}_p + \mathcal{O}_p^*$, then (as in [MFA1]), Riemann–Roch and Serre duality gives

$$1 = \dim H^0(\Sigma, \mathcal{O}_p K^{1/2}) - \dim H^0(\Sigma, \mathcal{O}_p^* K^{1/2})$$

and so $\varphi(V) = 1 = \varphi(\Omega)$.

Theorem 1. *Let S be a smooth curve in the total space of $\pi : K \to \Sigma$ given by an equation $x^n + a_1 x^{n-1} + \cdots + a_n = 0$ and let L be a line bundle on S such that $L^2 \cong K_S \pi^* K^*$. Define $V = \pi_* L$, the direct image bundle given the orthogonal structure described above.*

Let $K^{1/2}$ be a theta characteristic on Σ with $\varphi_\Sigma(1) = 0$, and $K_S^{1/2} = L\pi^* K^{1/2}$ the corresponding theta characteristic on S. Then

$$w_2(V) = \varphi_S(1) + \varphi_\Sigma(\alpha(w_1(V))).$$

Proof. The class of V in $KO(\Sigma)$ is $n - 1 + \alpha(w_1(V)) + w_2(V)\Omega$ so

$$w_2(V) = (n-1)\varphi_\Sigma(1) + \varphi_\Sigma(\alpha(w_1(V)) + \varphi_\Sigma(V)$$

and $\varphi_\Sigma(1) = 0$ by the choice of $K^{1/2}$.

Now, the defining property of the direct image is that $H^0(A, \pi_* L) = H^0(\pi^{-1}(A), L)$ for any open set $A \subset \Sigma$ so with $A = \Sigma$

$$H^0(S, K_S^{1/2}) = H^0(S, L\pi^* K^{1/2}) = H^0(\Sigma, \pi_* L \otimes K^{1/2})$$

which gives $\varphi_S(1) = \varphi_\Sigma(V)$. □

Remarks.

1. There is an alternative approach to deriving this formula using the topological definition of the mod 2 index due to Thurston (see [AGH] p. 291). Away from the branch locus B, the monodromy of a loop in $\Sigma \backslash B$ preserves the orthogonal structure on the direct image as a sum of squares and so lies in the group $B(n)$, the semi-direct product of the symmetric group $S(n)$ and $(\mathbf{Z}_2)^n$. This group is a subgroup of $O(n)$ and has a double covering $C(n) \subset Pin(n)$ which is a central extension by $\mathbf{Z}_2$. The authors of [EOP] relate the mod 2 invariant to lifting issues related to this group. In our context it clearly corresponds to the question of whether the structure group of V lifts to Spin(n), i.e., whether $w_2(V) = 0$ or not.

2. The characteristic class w_2 is independent of which spin structure $K^{1/2}$ we choose, (which was why it was convenient to take an even one in the theorem). A better way to formulate this fact is to regard the isomorphism $L^2 \cong K_S\pi^*K^*$ as a KO-theory orientation on the *map* $\pi : S \to \Sigma$. There is then a push-forward map $\pi_! : KO(S) \to KO(\Sigma)$ and $[V] = \pi_!(1)$.

Given the formula in Theorem 1 we need to determine how many points of order 2 give $w_2(V) = 0$. The interpretation of w_2 via φ_S tells us that this is an affine quadratic function ψ on P[2], the elements of order 2 in the Prym variety. Choosing an origin such that $\psi(0) = 0$ this means that $\psi(x + y) = \psi(x) + \psi(y) + (x, y)$ using the intersection form on $H^1(S, \mathbf{Z}_2)$ restricted to P[2].

The quadratic functions xy and $x^2 + xy + y^2$ on $(\mathbf{Z}_2)^2$ have the same bilinear form but are not equivalent and for quadratic functions associated with a non-degenerate bilinear form there are two canonical forms: a sum of k terms xy or a sum of $(k - 1)$ such terms plus a single term $x^2 + xy + y^2$. They are distinguished by their Arf invariant $\in \mathbf{Z}_2$ which is zero in the first case and 1 in the second. When the Arf invariant is 0, ψ has $2^{k-1}(2^k + 1)$ zeros and otherwise $2^{k-1}(2^k - 1)$. This interpretation shows that the invariant is independent of any choice of origin. The invariant is additive under orthogonal direct sum since $x^2+xy+y^2+u^2+uv+v^2 = (x + u)(x + y + u) + (y + v)(y + u + v)$.

Note in what follows that $\pi^* : H^1(\Sigma, \mathbf{Z}_2) \to H^1(S, \mathbf{Z}_2)$ is injective: indeed, (as in [BNR]), given a degree zero line bundle L on Σ with π^*L trivial, we have a section of $\pi_*\pi^*L = L \oplus LK^{-1} \oplus \dots$. Since L has degree 0 and $g > 1$ this must be a section of L which is therefore trivial.

Proposition 2. *In the context of Theorem 1, when n is odd there are* $2^{2p-1} + 2^{p-1}$ *choices of L which give* $w_1(V) = w_2(V) = 0$*; and when* $n = 2m$ *there are* $2^{2p-1} + (-1)^{m(g-1)}2^{p+g-1}$ *choices, where* $p = (g - 1)(n^2 - 1)$.

Proof. We have already observed that $w_1(V) = 0$ if U lies in the Prym variety P.

1. First consider the case where $n = 2m + 1$ is odd.

The Prym variety has polarization $(1, 1, 1, \dots, n, n, \dots n)$ with g copies of n (see [BNR]) so since n is odd the intersection matrix mod 2 is non-degenerate. Moreover, since $\mathrm{Nm}\,\pi^*(x) = (2m + 1)x = x$ if $2x = 0$, then $H^1(S, \mathbf{Z}_2)$ is an orthogonal direct sum $\pi^*H^1(\Sigma, \mathbf{Z}_2) \oplus$ P[2].

From Theorem 1 we have $w_2(V) = \varphi_S(1)$ as a function of theta characteristics of the form $K_S^{1/2} = LK^{1/2}$. By [MFA1] for all choices of $K_S^{1/2}$ the Arf invariant is 0. But the invariant is additive under orthogonal direct sum, and if we take $K_S^{1/2} = \pi^*(UK^{m+1/2})$ for $U \in H^1(\Sigma, \mathbf{Z}_2)$, then taking the direct image,

$$\begin{aligned}\dim H^0(S, K_S^{1/2}) &= \dim H^0(\Sigma, U \otimes (K^{-m+1/2} \oplus \dots \oplus K^{m+1/2}))\\ &= \dim H^0(\Sigma, U \otimes K^{1/2}) + 2(g - 1) + \dots + 2m(g - 1)\end{aligned}$$

which is $\dim H^0(\Sigma, U \otimes K^{1/2}) \bmod 2$. This is the standard quadratic function for the Riemann surface Σ so the Arf invariant is zero and by additivity so is the invariant on P[2].

We have $\dim H^0(\Sigma, K^{1/2}) = 0 \bmod 2$ by choice, the origin, and so it follows that there are $2^{p-1}(2^p + 1)$ zeros where $p = (g-1)\dim SL(n) = (g-1)(n^2-1)$ is the dimension of the Prym variety.

2. Now assume $n = 2m$.

In this case $\mathrm{Nm}\,\pi^*(x) = 2mx = 0$ and $\pi^* H^1(\Sigma, \mathbf{Z}_2)$ lies inside P[2], as the degeneracy subspace of the intersection form. In this case taking $K_S^{1/2} = \pi^*(UK^m)$ for $U \in H^1(\Sigma, \mathbf{Z}_2)$ we have

$$\begin{aligned}\dim H^0(S, K_S^{1/2}) &= \dim H^0(\Sigma, U \otimes (K^{-m+1} \oplus \cdots \oplus K^m)) \\ &= \dim H^0(\Sigma, U) + H^0(\Sigma, UK) + 3(g-1) + \dots \\ &\quad + (2m-1)(g-1)\end{aligned}$$

which by Riemann–Roch and Serre duality is $m(g-1) \bmod 2$.

From the proof of Theorem 1 in this case $w_2(V) = \varphi_S(1) + 2m\varphi_\Sigma(1) = \varphi_S(1)$ independently of the choice of $K^{1/2}$. In particular this means that $\varphi_S|_{\mathrm{P}[2]}$ is invariant under the action of $U \in \pi^* H^1(\Sigma, \mathbf{Z}_2)$ and hence for $y \in$ P[2] and $x \in \pi^* H^1(\Sigma, \mathbf{Z}_2)$, $\psi(x+y) = \psi(y)$, and so $\psi(x) = 0$ and $(x, y) = 0$.

Choose a transverse $2p - 2g$-dimensional subspace X to $\pi^* H^1(\Sigma, \mathbf{Z}_2)$ and consider the quadratic function ψ restricted to X. Then from the canonical form there is a basis y_i, z_i of X such that the function is $\sum_{i=1}^{p-g} a_i b_i$ or $\sum_{i=2}^{p-g} a_i b_i + a_1^2 + a_1 b_1 + b_1^2$.

Take a basis $x_1, \dots, x_{2g}$ for $\pi^* H^1(\Sigma, \mathbf{Z}_2)$, then by non-degeneracy of the intersection form on $H^1(S, \mathbf{Z}_2)$ there are elements w_i such that $(x_i, w_j) = \delta_{ij}$.

Take

$$\tilde{w}_1 = w_1 + \sum_i (w_1, y_i) z_i + \sum_i (w_1, z_i) y_i$$

which makes $\tilde{w}_1$ orthogonal to X. Since each x_i is orthogonal to P[2] the 2-dimensional space spanned by $x_1, \tilde{w}_1$ is orthogonal to X. Then since $\psi(x_1) = 0$

$$\psi(a\tilde{w}_1 + bx_1) = a^2\psi(\tilde{w}_1) + b^2\psi(x_1) + ab = a(a\psi(\tilde{w}_1) + b)$$

and so on the space spanned by X and these two vectors we are adding an xy term, which means we have the same Arf invariant. By induction so does the full space.

Now, as we showed above, for $U \in \pi^* H^1(\Sigma, \mathbf{Z}_2)$ we have $\dim H^0(S, K_S^{1/2}) = m(g-1) \bmod 2$, so the number of zeros on X is $2^{p-g-1}(2^{p-g} + (-1)^{m(g-1)})$. Acting by $U \in \pi^* H^1(\Sigma, \mathbf{Z}_2)$ gives all of P[2] and hence the total number of zeros is

$$2^{2g} \times 2^{p-g-1}(2^{p-g} + (-1)^{m(g-1)}) = 2^{2p-1} + (-1)^{m(g-1)} 2^{p+g-1} \tag{6}$$

□

6 Characteristic Classes for $Sp(2m, \mathbf{R})$

We have already seen in Sect. 4 that a Higgs bundle for the group $Sp(2m,\mathbf{R})$ is obtained from the spectral curve S defined by $x^{2m}+a_2x^{2m-2}+\cdots+a_{2m}=0$ as the direct image $V=W\oplus W^*$ of a line bundle $L\cong U\pi^*K^{(n-1)/2}$ such that $\sigma^*U\cong U$ and U^2 is trivial. Moreover W is the direct image of the invariant sections of L. The characteristic class here is $c_1(W)$, and we need to evaluate this as a function of P[2] for the Prym variety of $p: S\to S/\sigma=\bar{S}$.

Note first that, as in the case of $SL(2m,\mathbf{R})$, $p^*H^1(\bar{S},\mathbf{Z}_2)$ lies in the Prym variety. The map p^* is also injective. This is a similar argument to the one above. In this case the invariant and anti-invariant parts decompose the direct image of p^*U as $U\oplus U'$ where U' has negative degree. So if p^*U has a section, so does U. The dimension of the $\mathbf{Z}_2$-vector space P[2]$/p^*H^1(\bar{S},\mathbf{Z}_2)$ is therefore $2(g_S-2g_{\bar{S}})$, but by Riemann–Hurwitz, since $S\to\bar{S}$ has $4m(g-1)$ branch points, $2-2g_S=2(2-2g_{\bar{S}})-4m(g-1)$ and so

$$\dim P[2]/p^*H^1(\bar{S},\mathbf{Z}_2)=4m(g-1)-2.$$

We now use the condition $\sigma^*U\cong U$, so that the involution lifts to the line bundle U. There are two lifts $\pm\sigma$ but fix attention on one for the moment. Following [LS1], we consider the action ±1 of σ on the fibre of U at a fixed point.

Proposition 1. *Suppose the action is -1 at ℓ fixed points, then $c_1(W)=-\ell/2+m(g-1)$.*

Proof. The fixed point set of σ is the intersection of the zero section of K with S. Setting $x=0$ in the equation $x^{2m}+a_2x^{2m-2}+\cdots+a_{2m}=0$, these points are the images of the $4m(g-1)$ zeros of $a_m\in H^0(\Sigma,K^{2m})$ under the zero section. The action is -1 at ℓ of these points.

Choose a line bundle M on Σ of large enough degree that the higher cohomology groups vanish and then applying the holomorphic Lefschetz formula [AB] we obtain

$$\dim H^0(S,L\pi^*M)^+-\dim H^0(S,L\pi^*M)^-=\frac{1}{2}(-\ell+(4m(g-1)-\ell))$$

where the superscript denotes the ±1 eigenspace under the action of σ. Riemann–Roch gives

$$\dim H^0(S,L\pi^*M)^++\dim H^0(S,L\pi^*M)^-=\dim H^0(\Sigma,V\otimes M)=2m(1-g+c_1(M))$$

since V is symplectic and $\deg V=0$. Hence

$$\dim H^0(S,L\otimes\pi^*M)^+=-\frac{\ell}{2}+mc_1(M)$$

But by the definition of W this is $\dim H^0(\Sigma, W \otimes M)$ and Riemann–Roch and the vanishing of H^1 give the value $m(1-g) + c_1(W) + mc_1(M)$ and so

$$c_1(W) = -\frac{\ell}{2} + m(g-1). \tag{7}$$

□

Remarks.

1. Since $0 \leq \ell \leq 4m(g-1)$ we have $|c_1(W)| \leq m(g-1)$ which is the Milnor–Wood inequality for this group [PG, GGM].
2. Taking the action $-\sigma$ instead of σ changes ℓ in the formula to $4m(g-1) - \ell$ and $c_1(W)$ to $-c_1(W)$, and the roles of W and W^* are interchanged. Choosing one or the other is simply a choice of generator in $\pi_1(U(m)) = \mathbf{Z}$.

The formula (7) above clearly requires ℓ to be even, but there is a reason for this. If $\ell = 0$ and the action is trivial at all fixed points, then U is the pull-back of a flat line bundle of order 2 on the quotient $\bar{S}$. In general, let B denote the subset of ℓ points in the branch locus of $S \to \bar{S}$, then the line bundle corresponds to a flat line bundle on $\bar{S}\backslash B$ where the local holonomy around each $b \in B$ is -1. The global holonomy defines a homomorphism $\rho : \pi_1(\Sigma\backslash B) \to \mathbf{Z}_2$ where the fundamental group has generators A_i, B_i, $1 \leq i \leq g$ and δ_j, $1 \leq j \leq N$, each δ_i defining a loop around a branch point. These satisfy the relation

$$\prod_i [A_i, B_i] \prod_j \delta_j = 1$$

but then, with values in the abelian group $\mathbf{Z}_2$ we must have $\prod_j \delta_j = 1$ and hence an even number of -1 terms.

This interpretation helps to understand which of the 2^{2p} (where $p = \dim \mathrm{P}(S, \Sigma)$) elements in the Prym variety yield a given characteristic class. Let Z be the $4m(g-1)$-element set of zeros of the section a_{2m} of K^{2m} and let $C(Z)$ be the space of $\mathbf{Z}_2$-valued functions on Z, and $C_0(Z)$ the subspace of those whose integral is zero, i.e., takes the value 1 an even number of times. The constant function 1 lies in $C_0(Z)$ and let $H(Z)$ be the $(4m(g-1)-2)$-dimensional quotient.

For a line bundle $U \in \mathrm{P}[2]$ let A be the subset of Z over which the action of σ is -1. (The set $A \subset \Sigma$ is in bijection with $B \subset \bar{S}$ since B lies in the zero section of K^m.) As noted above, A has an even number of elements and so its characteristic function χ_A lies in $C_0(Z)$. Define $f(U) \in H(Z)$ to be its equivalence class. Since we take the quotient by the constant function 1, this is independent of the choice of lift of σ.

Proposition 2. *The homomorphism f from* $\mathrm{P}[2]$ *to $H(Z)$ is surjective and has kernel* $p^*H^1(\bar{S}, \mathbf{Z}_2) \subset \mathrm{P}[2]$.

Proof. Given *any* subset of the zeros of a_{2m} with an even number of elements we can choose $\delta_i = -1$ as above and get a flat line bundle $U \in H^1(S, \mathbf{Z}_2)$ and an action of σ which acts as -1 at those points, so the homomorphism is surjective. The kernel consists of line bundles with trivial action at all fixed points and these are precisely those pulled back from the quotient $\bar{S}$. □

Since $H(Z)$ and $\mathrm{P}[2]/p^*H^1(\bar{S}, \mathbf{Z}_2)$ have the same number of elements we see from the proposition that they are isomorphic.

We can now count the points in P[2] with fixed characteristic class. This is, from Eq. (7), $c_1(W) = m(g-1) - \ell/2$ and involves a choice between W and W^* or equivalently a choice of lifting of the involution σ, or the subset $A \subset Z$ with an even number ℓ of elements. Thus the number of such elements in P[2] is, from Proposition 2,

$$\binom{4m(g-1)}{\ell} \times 2^{2q},$$

where $q = g_{\bar{S}} = (2m^2 - m)(g-1) + 1$.

Remarks.

1. The $\mathbf{Z}_2$-vector space $H(Z)$ is a representation of the symmetric group $S(4m(g-1))$, the permutations of the $4m(g-1)$ branch points, and we may describe the above result by saying that the characteristic class is determined by the orbit of the symmetric group on this space. In the case of $n = 2$ this picture was derived via the monodromy action of the family of abelian varieties in [LS0].
2. For $n = 2$ the two groups coincide, so we may use the formula above to compare with the $SL(2, \mathbf{R})$ case. Here S is a double cover of Σ and so $\bar{S} = \Sigma$, hence $H(Z) = \mathrm{P}[2]/\pi^*H^1(\Sigma, \mathbf{Z}_2)$. Equation (7) gives $c_1(W) = -\ell/2 + (g-1)$ and $w_2 = c_1(W) \bmod 2$, but here we don't distinguish between W and W^* so the number with $w_2 = 0$ is

$$\frac{1}{2} \sum_{\ell \equiv (2g-2) \bmod 4} \binom{4(g-1)}{\ell} \times 2^{2g}.$$

If $(g-1)$ is even this is

$$2^{2g-3}((1+1)^{4(g-1)} + (1+i)^{4(g-1)} + (1-1)^{4(g-1)} + (1-i)^{4(g-1)})$$

and if $(g-1)$ is odd

$$2^{2g-3}((1+1)^{4(g-1)} + (1+i)^{4(g-1)} - (1-1)^{4(g-1)} - (1-i)^{4(g-1)}).$$

Using $e^{i\pi/4} = (1+i)/\sqrt{2}$ this gives $2^{6g-7} + 2^{4g-4}$ or $2^{6g-7} - 2^{4g-4}$ which checks with (6).

7 Mirror Symmetry

A hyperkähler manifold has not only three complex structures I, J, K but also three symplectic forms, the corresponding Kähler forms $\omega_1, \omega_2, \omega_3$. A *brane* for a complex manifold (a B-brane) is roughly speaking a holomorphic bundle over a complex submanifold and for a symplectic manifold (an A-brane) it is a flat vector bundle over a Lagrangian submanifold. For a hyperkähler manifold a BAA-brane is a B-brane for the complex structure I and an A-brane for the symplectic structures ω_2, ω_3. The trivial bundle over a component of the G^r-character variety is an example. As the fixed point set of an antiholomorphic isometry for complex structures J, K it is Lagrangian for the Kähler forms ω_2, ω_3, and it is holomorphic with respect to I.

Mirror symmetry should transform a BAA-brane on a hyperkähler manifold M to a BBB-brane on its mirror $\hat{M}$ (I am indebted to Sergei Gukov for this information). A BBB-brane is a holomorphic bundle over a complex submanifold with respect to all complex structures I, J and K, equivalently a hyperkähler submanifold with a hyperholomorphic bundle over it. A hyperholomorphic bundle is a bundle with connection whose curvature is of type $(1,1)$ with respect to all complex structures. Such connections (generalizations of instantons in four dimensions) are quite rare and so it is intriguing to seek such an object as the mirror of a G^r-character variety. We shall attempt this now for the group $G^r = U(m,m) \subset GL(2m, \mathbf{C})$.

The Higgs bundle and spectral data description of $U(m,m)$ will be familiar from our previous discussion of $Sp(2m, \mathbf{R})$. Details can be found in [LS1]. The Higgs bundle is of the form $V = W_1 \oplus W_2$ and the Higgs field Φ is off-diagonal: $\Phi(w_1, w_2) = (\beta(w_2), \gamma(w_1))$. There is a characteristic class $c_1(W_1)$, and to keep the link with flat connections we need $c_1(V) = 0$ and so $c_1(W_1) = -c_1(W_2) \in \mathbf{Z}$. The spectral curve S has the form $x^{2m} + a_2 x^{2m-2} + \cdots + a_{2m} = 0$ and hence an involution $\sigma(x) = -x$ and the spectral data consist of taking a line bundle L on S such that $\sigma^* L \cong L$. As in Sect. 6, the characteristic class is determined by the number of points on $x = 0$ at which the lifted action of σ is -1. The difference here with the $Sp(2m, \mathbf{R})$-case is that the fibre is not discrete but is instead the disjoint union of a finite number of abelian varieties. In fact if L_1, L_2 are two line bundles with the same subset of fixed points at which σ acts as -1, then the action on $L_1^* L_2$ is trivial and so it is pulled back from the quotient $\bar{S}$. Thus the fibre is isomorphic to the disjoint union of N copies of $\mathrm{Pic}^0(\bar{S})$ where $N = 2^{4m(g-1)-1}$ is the number of subsets of the zero set Z of a_{2m} with an even number of elements.

For a Calabi–Yau manifold with a special Lagrangian fibration mirror symmetry is effected via the Strominger–Yau–Zaslow approach of replacing each nonsingular torus fibre by its dual, and hoping that it can be extended over the discriminant locus in the base. The Higgs bundle integrable system fits into this framework as first investigated in [HT]. As in [Hit4] for certain cases and [DP] in general, it corresponds to replacing the group G^c by its Langlands dual group ${}^L G^c$.

We now consider the structure on the dual fibration relevant for $U(m,m) \subset GL(2m, \mathbf{C})$. The abelian variety for $GL(2m, \mathbf{C})$ is the Jacobian, or $\mathrm{Pic}^0(S)$, and since

Jacobians are self-dual, and the norm map is the adjoint of the pull-back, dualizing the inclusion $\mathrm{Pic}^0(\bar{S}) \subset \mathrm{Pic}^0(S)$ gives

$$0 \to \mathrm{P}(S, \bar{S}) \to \mathrm{Pic}^0(S) \to \mathrm{Pic}^0(\bar{S}) \to 0$$

and the Prym variety $\mathrm{P}(S, \bar{S})$ is a distinguished subvariety of $\mathrm{Pic}^0(S)$. In terms of duality it parametrizes line bundles on $\mathrm{Pic}^0(S)$ which are trivial on $\mathrm{Pic}^0(\bar{S})$.

If mirror symmetry is to work as predicted this family of abelian varieties over the space of polynomials $x^{2m} + a_2 x^{2m-2} + \cdots + a_{2m}$ should extend to a hyperkähler submanifold of the Higgs bundle moduli space for the Langlands dual of $GL(2m, \mathbf{C})$, which is again $GL(2m, \mathbf{C})$. But in Sect. 4 we saw that the $Sp(2m, \mathbf{C})$-moduli space had an integrable system over this base whose generic fibre was $\mathrm{P}(S, \bar{S})$. The inclusion of a group gives a hyperkähler subspace of the moduli space so the symplectic Higgs bundles form a hyperkähler subspace of the moduli space of $GL(2m, \mathbf{C})$-Higgs bundles. We therefore have the first requirement of the mirror—the hyperkähler support for a hyperholomorphic bundle.

The remaining task is to find a hyperholomorphic vector bundle over the $Sp(2m, \mathbf{C})$-moduli space, or rather several, one for each characteristic class. There are relatively few constructions of such bundles but there is one which involves a Dirac-type operator, and which we describe next. More information may be found in [Hit5, Bon].

For each (A, Φ) satisfying the Higgs bundle equations for a compact group G, we take a vector bundle V associated with the principal G-bundle via a representation of G and define an elliptic operator $D^* : V \otimes (K \oplus \bar{K}) \to V \otimes (K\bar{K} \oplus K\bar{K})$ by

$$D^* = \begin{pmatrix} \bar{\partial}_A & \Phi \\ \Phi^* & \partial_A \end{pmatrix}.$$

The equation $F_A + [\Phi, \Phi^*] = 0$ yields a vanishing theorem for irreducible connections and the index theorem gives $\dim \ker D^* = (2g-2)\,\mathrm{rk}\, V$. For the adjoint representation the null-space can be viewed as the tangent space of the moduli space.

Let $\mathcal{M}$ be the Higgs bundle moduli space for a linear group—the moduli space of pairs (V, Φ) with possible extra structure. Given a universal bundle over $\mathcal{M} \times \Sigma$, the family of null-spaces for D^* defines a rank $(2g-2)\,\mathrm{rk}\, V$ vector bundle on $\mathcal{M}$, and since D^* acts on one-forms with values in a Hermitian bundle V, there is a conformally invariant $\mathcal{L}^2$ inner product which defines by projection a connection on this bundle. It turns out that this connection is hyperholomorphic: for the adjoint representation it is the Levi–Civita connection. We shall leave till later the issue of the existence of a universal bundle—locally these exist and connections are locally determined.

The null-space of D^* can be viewed in different ways according to the complex structures I, J, K. For J the operator D^* is the Hodge operator for the de Rham complex of the flat connection $\nabla_A + \Phi + \Phi^*$; for I it is the Hodge operator for the total differential $\bar{\partial} \pm \Phi$ in the double complex

$$\Omega^{0,*}(V) \stackrel{\Phi}{\to} \Omega^{0,*}(V).$$

In this latter case $\ker D^*$ is identified with the hypercohomology group $\mathbf{H}^1$. Trading the Dolbeault viewpoint for the Čech approach, it is the hypercohomology for the complex of sheaves

$$\mathcal{O}(V) \stackrel{\Phi}{\to} \mathcal{O}(V \otimes K).$$

Now take V to be associated with the defining representation of $Sp(m)$, and (A, Φ) to lie in a generic fibre of the integrable system. Then $\Phi : V \to V \otimes K$ is generically an isomorphism and the spectral sequence for the complex of sheaves identifies the hypercohomology group $\mathbf{H}^1$ with sections of a sheaf supported on the zero set of $\det \Phi$. For a smooth spectral curve, $\det \Phi = a_{2m}$ has $4m(g-1)$ distinct zeros Z and we have

$$\mathbf{H}^1 \cong \bigoplus_{z \in Z} \operatorname{coker} \Phi_z.$$

From the spectral data, $L\pi^*K = \operatorname{coker}(x - \Phi)$, so identifying Z with the intersection of the spectral curve S with the zero section $x = 0$ of K we have

$$\mathbf{H}^1 \cong \bigoplus_{z \in Z} (L\pi^*K)_z$$

(note in particular the dimension checks with the index theory calculation).

To summarize, what we have here is a hyperholomorphic bundle $\mathbf{V}$ over the $Sp(2m, \mathbf{C})$-moduli space whose fibre at a point defined by a nonsingular spectral curve S and line bundle L is given by $\oplus_{z \in Z} (L\pi^*K)_z$.

Now observe that the components of the fibre in the $U(m,m)$-character variety with a fixed characteristic class correspond to the subsets of $\ell = 2k$ elements in Z, and, at a point in the $Sp(2m, \mathbf{C})$-moduli space over the same point in the base

$$\Lambda^{2k}\mathbf{V} \cong \bigoplus_{\{z_1,\ldots,z_{2k}\} \subset Z} (L\pi^*K)_{z_1} (L\pi^*K)_{z_2} \cdots (L\pi^*K)_{z_{2k}}$$

is a sum over all such subsets. This bundle, with its induced hyperholomorphic connection, seems a natural choice for the mirror: as a direct sum over components of the fibre it is analogous to the Fourier–Mukai transform yet it is well-defined on the whole moduli space apart from the issue of the universal bundle, which we consider next.

The $Sp(2m, \mathbf{C})$-moduli space has no universal bundle: the obstruction lies in $H^2(\mathcal{M}, \mathbf{Z}_2)$. To be more concrete, for an open covering $\{U_\alpha\}$ of $\mathcal{M}$ there is a local universal bundle $\mathcal{V}_\alpha$ on $U_\alpha \times \Sigma$ and on $U_\alpha \cap U_\beta$ there is a line bundle $\mathcal{L}_{\alpha\beta}$ of order 2 such that $\mathcal{V}_\beta \cong \mathcal{V}_\alpha \otimes \mathcal{L}_{\alpha\beta}$ with compatibility conditions on the isomorphisms. This describes the gerbe which is the obstruction. The D^* null-spaces define bundles $\mathbf{V}_\alpha$ over U_α with a hyperholomorphic connection and these are related on the intersection $U_\alpha \cap U_\beta$ by the flat line bundle $\mathcal{L}_{\alpha\beta}$. However, the even exterior power $\Lambda^{2k}\mathbf{V}_\alpha$ is insensitive to this ambiguity and so is well-defined globally.

Remarks.

1. The operator D^* may be regarded as a quaternionic operator (see [Hit5]) with coefficients in the bundle V whose structure group is $Sp(m)$ hence also quaternionic. This means the null-space has a real structure and a Hermitian structure, equivalently an orthogonal structure. In particular we have an isomorphism as bundles with connection $\Lambda^{2k}\mathbf{V} \cong \Lambda^{4m(g-1)-2k}\mathbf{V}$. As we have seen it is only the choice of lifting of the involution σ that distinguishes a $2k$-element subset of Z and its complement, so this is expected.
2. From the point of view of the spectral data there is a natural orthogonal structure which is almost certainly the same as the above differential-geometric description. At each point $z \in Z$ we have $a_{2m}(z) = 0$ and the derivative at a simple zero defines $da_{2m}(z) \in K_z^{2m+1}$. But $L \cong U\pi^*K^{(2m-1)/2}$ and $\sigma^*U \cong U^*$, so at the fixed point z of σ we have a non-zero vector u_z in U_z^{-2}. Given $s \in (L\pi^*K)_z$ we can define $s^2u_z/da_{2m}(z) \in \mathbf{C}$ and summing these get a non-degenerate quadratic form on $\mathbf{V}$.
3. The cases $k = 0$ or $k = 2m(g-1)$ correspond to the maximum absolute value of the characteristic class allowed and here the hyperholomorphic bundle is the trivial line bundle $\Lambda^0\mathbf{V}_\alpha$. Maximal representations play a special role in the study of character varieties (see, e.g., [Burg]).

References

[AB] M.F. Atiyah, R. Bott, A Lefschetz fixed point formula for elliptic complexes II. Applications. Ann. Math. **88**, 451–491 (1968)

[AGH] E. Arbarello, M. Cornalba, P.A. Griffiths, J. Harris, *Geometry of Algebraic Curves*, vol. I (Springer, New York, 1985)

[BNR] A. Beauville, M.S. Narasimhan, S. Ramanan, Spectral curves and the generalised theta divisor. J. Reine Angew. Math. **398**, 169–179 (1989)

[Bon] J. Bonsdorff, Autodual connection in the Fourier transform of a Higgs bundle. Asian J. Math. **14**, 153–173 (2010)

[Burg] M. Burger, A. Iozzi, A. Wienhard, Surface group representations with maximal Toledo invariant. Ann. Math. **172**, 517–566 (2010)

[DP] R. Donagi, T. Pantev, Langlands duality for Hitchin systems. Invent. Math. **189**, 653–735 (2012)

[EOP] A. Eskin, A. Okounkov, R. Pandharipande, The theta characteristic of a branched covering. Adv. Math. **217**, 873–888 (2008)

[GGM] O. Garcia-Prada, P. Gothen, I. Mundet i Riera, Higgs bundles and surface group representations in the real symplectic group. J. Topol. **6**, 64–118 (2013)

[Hirz1] F. Hirzebruch, Problems on differentiable and complex manifolds. Ann. Math. **60**, 213–236 (1954)

[Hit1] N.J. Hitchin, The self-duality equations on a Riemann surface. Proc. Lond. Math. Soc. **55**, 59–126 (1987)

[Hit2] N.J. Hitchin, Stable bundles and integrable systems. Duke Math. J. **54**, 91–114 (1987)

[Hit3] N.J. Hitchin, Lie groups and Teichmüller space. Topology **31**, 449–473 (1992)

[Hit4] N.J. Hitchin, Langlands duality and G_2 spectral curves. Q. J. Math. Oxf. **58**, 319–344 (2007)

[Hit5] N.J. Hitchin, The Dirac operator, in *Invitations to Geometry and Topology*, ed. by M. Bridson, S. Salamon. Oxford Graduate Texts in Mathematics (Oxford University Press, Oxford, 2002), pp. 208–232

[HT] T. Hausel, M. Thaddeus, Mirror symmetry, Langlands duality and Hitchin systems. Invent. Math. **153**, 197–229 (2003)

[LS] L.P. Schaposnik, Spectral data for G-Higgs bundles. D. Phil Thesis, Oxford (2013)

[LS0] L.P. Schaposnik, Monodromy of the SL_2 Hitchin fibration. Int. J. Math. **24**, 1350013 (2013)

[LS1] L.P. Schaposnik, Spectral data for $U(m,m)$-Higgs bundles. Int. Math. Res. Not. (2014). doi:10.1093/imrn/rnu029

[MFA1] M.F. Atiyah, Riemann surfaces and spin structures. Ann. Sci. École Norm. Sup. **4**, 47–62 (1971)

[PG] P. Gothen, The topology of Higgs bundle moduli spaces. Ph.D. thesis, Warwick (1995)

[Sim] C. Simpson, Higgs bundles and local systems. Inst. Hautes Études Sci. Publ. Math. **75**, 5–95 (1992)

[Tr] W. Trench, On the eigenvalue problem for Toeplitz band matrices. Linear Algebra Appl. **64**, 199–214 (1985)

Hirzebruch–Milnor Classes and Steenbrink Spectra of Certain Projective Hypersurfaces

Laurentiu Maxim, Morihiko Saito, and Jörg Schürmann

To the Memory of Friedrich Hirzebruch

Abstract We show that the Hirzebruch–Milnor class of a projective hypersurface, which gives the difference between the Hirzebruch class and the virtual one, can be calculated by using the Steenbrink spectra of local defining functions of the hypersurface if certain good conditions are satisfied, e.g., in the case of projective hyperplane arrangements, where we can give a more explicit formula. This is a natural continuation of our previous paper on the Hirzebruch–Milnor classes of complete intersections.

1 Introduction

In his classical book [Hi], Hirzebruch introduced the cohomology Hirzebruch characteristic class $T_y^*(TX)$ of the tangent bundle TX of a compact complex manifold X, see also Sect. 2.1 below. It belongs to $\mathbf{H}^\bullet(X)[y]$ where $\mathbf{H}^k(X) = H^{2k}(X, \mathbf{Q})$. By specializing to $y = -1, 0, 1$, it specializes to the Chern class $c^*(TX)$, the Todd class $td^*(TX)$, and the Thom–Hirzebruch L-class $L^*(TX)$, respectively, see [HiBeJu, Sect. 5.4]. Its highest degree part, which is called the T_y-genus in [Hi, 10.2], was mainly interested there by the relation with his Riemann–Roch theorem. This coincides with the χ_y-genus

L. Maxim
Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Drive, Madison, WI 53706-1388, USA
e-mail: maxim@math.wisc.edu

M. Saito (✉)
RIMS Kyoto University, Kyoto 606–8502, Japan
e-mail: msaito@kurims.kyoto-u.ac.jp

J. Schürmann
Mathematisches Institut, Universität Münster, Einsteinstraße 62, 48149 Münster, Germany
e-mail: jschuerm@uni-muenster.de

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_9

$$\chi_y(X) := \sum_p \chi(\Omega_X^p)\, y^p \in \mathbf{Z}[y],$$

which specializes to the Euler characteristic, the arithmetic genus, and the signature of X for $y = -1, 0, 1$. The cohomology class $T_y^*(TX)$ is identified by Poincaré duality with the homology Hirzebruch class $T_{y*}(X)$ in the smooth case. It is generalized to the singular case by [BrScYo] (see Sect. 2.2 below).

Hirzebruch also introduced there the virtual T_y-genus (or T_y-characteristic) which gives the T_y-genus of smooth complete intersections X in smooth projective varieties Y. We can define the virtual Hirzebruch class $T_{y*}^{\mathrm{vir}}(X)$ of any complete intersection X like the virtual T_y-genus even if X is singular. In this paper we adopt a more sophisticated construction as in [BrScYo], see Sects. 2.2–2.3 below. This is compatible with the construction in [Hi, 11.2], see Sect. 2.4 below. In [MaSaSc1] we proved that the difference between the Hirzebruch class and the virtual one is given by the *Hirzebruch–Milnor class* $M_y(X)$ supported on the singular locus of X, and gave an inductive formula in the case of global complete intersections with arbitrary singularities.

In this paper, we restrict to the case of projective hypersurfaces (i.e., the codimension is one) satisfying certain good conditions in order to prove a formula for the Hirzebruch–Milnor class $M_y(X)$ by using the Steenbrink spectra (see [St1, St2]) of local defining functions of X in Y. This is a natural continuation of the last section of [MaSaSc1]. (Note that the implication of the calculations in loc. cit. was not explained there.) More precisely, let Y be a smooth complex projective variety having a very ample line bundle L. Set

$$X = s^{-1}(0) \quad \text{with } \; s \in \Gamma(Y, L^{\otimes m}) \; \text{ for some positive integer } \; m.$$

Let $s'_1, \dots, s'_{n+1}$ be sufficiently general sections of L, where $n := \dim X$. Take sufficiently general nonzero complex numbers a_j with $|a_j|$ sufficiently small ($j \in [1, n]$). For $j \in [1, n+1]$, set

$$s_{a,j} := s - a_1 s_1'^{\,m} - \cdots - a_{j-1} s_{j-1}'^{\,m}, \quad f_{a,j} := \big(s_{a,j}/s_j'^{\,m}\big)|_{Y\setminus X'_j}, \quad X'_j := s_j'^{-1}(0),$$

$$X_{a,j} := s_{a,j}^{-1}(0), \quad \Sigma_j := \mathrm{Sing}\, X_{a,j} \left(= \bigcap_{k<j} X'_k \cap \Sigma \right), \quad \Sigma := \Sigma_1.$$

Set $r := \max\{j \mid \Sigma_j \neq \emptyset\}$. By [MaSaSc1, MaSaSc2], there is the Hirzebruch–Milnor class $M_y(X) \in \mathbf{H}_\bullet(\Sigma)[y]$ with $\mathbf{H}_k(\Sigma) = H_{2k}^{\mathrm{BM}}(\Sigma, \mathbf{Q})$ or $\mathrm{CH}_k(\Sigma)_{\mathbf{Q}}$, satisfying

$$T_{y*}^{\mathrm{vir}}(X) - T_{y*}(X) = (i_{\Sigma,X})_* M_y(X), \tag{1}$$

$$M_y(X) = \sum_{j=1}^{r} T_{y*}\big((i_{\Sigma_j\setminus X'_j,\Sigma})_!\, \varphi_{f_{a,j}} \mathbf{Q}_{h,Y\setminus X'_j}\big), \tag{2}$$

where $i_{A,B} : A \hookrightarrow B$ denotes the inclusion for $A \subset B$ in general. In this paper, $\varphi_{f_{a,j}}\mathbf{Q}_{h,Y\setminus X'_j}$ denotes a mixed Hodge module up to a shift of complex on $\Sigma_j \setminus X'_j$ such that its underlying $\mathbf{Q}$-complex is the vanishing cycles $\varphi_{f_{a,j}}\mathbf{Q}_{Y\setminus X'_j}$ in [De3], see [Sa1, Sa2]. For the definition of $T_{y*}(\mathcal{M}^\bullet)$ with $\mathcal{M}^\bullet$ a bounded complex of mixed Hodge modules, see (12) below. In [MaSaSc1] we assumed $m = 1$, but it is not difficult to generalize the argument there to the case $m > 1$ by using Sect. 3.4 below, see [MaSaSc2] for details. By [Sch, Proposition 5.21], this formula specializes at $y = -1$ to a formula for the Chern classes, which was conjectured by Yokura [Yo2], and was proved by Parusiński and Pragacz [PaPr] (where $m = 1$).

In the case of hyperplane arrangements, we may assume

$$f_{a,j} = g_j + x_1^m + \cdots + x_{j-1}^m,$$

for some coordinates $x_1, \ldots, x_{n+1}$ of $Y \setminus X'_j = \mathbf{C}^{n+1}$, where g_j defines the restriction of the hyperplane arrangement to $Y \setminus X'_j$. We have a topologically trivial one-parameter family

$$g_j(\lambda x_1, \ldots, \lambda x_{j-1}, x_j, \ldots, x_{n+1}) + x_1^m + \cdots + x_{j-1}^m \quad (\lambda \in \mathbf{C}),$$

and apply the Thom–Sebastiani theorem [Sa3] at $\lambda = 0$ (together with [DiMaSaTo]). This argument can be extended to the general case by using the deformation to the normal cone. Set $Y_j := \bigcup_{k<j} X'_k$, $g_{a,j} := f_{a,j}|_{Y_j\setminus X'_j}$. For the calculation of the right-hand side of (2), it is then sufficient to calculate $\varphi_{g_{a,j}}\mathbf{Q}_{h,Y_j\setminus X'_j}$ together with the action of the semisimple part of the monodromy T_s, see [MaSaSc2] for details. From now on, we fix $j \in [1, r]$, and denote $g_{a,j}$, Y_j, $Y_j \cap X'_j$, Σ_j, respectively, by f, Y, X', Σ to simplify the notation.

Let $\mathcal{S}$ be a complex algebraic stratification of $\Sigma \setminus X'$ such that the $\mathcal{H}^j_S := \mathcal{H}^j\varphi_f\mathbf{Q}_{Y\setminus X'}|_S$ are local systems for any strata $S \in \mathcal{S}$ (which are assumed smooth). These local systems canonically underlie admissible variations of mixed Hodge structure $\mathbf{H}^j_S$, since $\varphi_f\mathbf{C}_{Y\setminus X'}$ underlies a mixed Hodge module up to a shift of complex $\varphi_{h,f}\mathbf{Q}_{Y\setminus X'}$. Let $\mathcal{H}^j_{S,\lambda} \subset \mathcal{H}^j_S$ be the λ-eigenspace by the action of the semisimple part T_s of the Milnor monodromy T (which is defined as the monodromy of the local system on a punctured disk associated with the Milnor fibration, see [De3]). The local system monodromy of $\mathcal{H}^j_{S,x}$ around X' coincides with the mth power of the Milnor monodromy where we take $x \in S$ sufficiently near X' so that we have a loop around X' passing through x. (This can be reduced to Lemma 4.3 below by using the expression $f = (s/s'^m)|_{Y\setminus X'}$ together with the deformation to the normal cone of X'.) So the situation is quite different from the one in [CaMaScSh] where X is a fiber of a morphism to a curve, and a formula for the Hirzebruch–Milnor class is given by using the mixed Hodge structure on each stalk of the vanishing cycles under the assumption on the triviality of the global monodromies of the local systems.

In this paper we prove a variant of it by using the Steenbrink spectrum (see Sect. 2.5 below):

$$\mathrm{Sp}(f,x) = \sum_{\alpha\in\mathbf{Q}} n_{f,x,\alpha} t^{\alpha} \quad (x \in \Sigma \setminus X').$$

Here $n_{f,x,\alpha} \in \mathbf{Z}$ is independent of $x \in S$, and will be denoted by $n_{f,S,\alpha}$. We say that a compactification $\tilde{S}$ of S is *good* if $D_{\tilde{S}} := \tilde{S} \setminus S$ is a divisor with simple normal crossings on a smooth projective variety $\tilde{S}$ and the natural inclusion $S \hookrightarrow \Sigma$ extends to a (unique) morphism $\pi_{\tilde{S}} : \tilde{S} \to \Sigma$. Using the results in the last section of [MaSaSc1], we get the following.

Theorem 1.1. *Assume the following two conditions:*

(*a*) *Every $\mathbf{H}^j_S$ is a locally constant variation of mixed Hodge structure on S.*
(*b*) *Each $\mathcal{H}^j_{S,\lambda}$ is isomorphic to a direct sum of copies of a rank 1 local system $L_{S,\lambda}$ which is independent of j.*

Let $\mathcal{L}_{\tilde{S},\lambda}$ be the Deligne extension of $L_{S,\lambda}$ as an $\mathcal{O}_{\tilde{S}}$-module with a logarithmic connection such that the eigenvalues of the residues of the connection are contained in $(0,1]$ where $\tilde{S}$ is any good compactification of S. Then

$$T_{y*}\big((i_{\Sigma\setminus X',\Sigma})_! \varphi_f \mathbf{Q}_{h,Y\setminus X'}\big) = \sum_{S,\alpha,q} (-1)^{q+n-1} n_{f,S,\alpha} (\pi_{\tilde{S}})_* td_{(1+y)*}\big[\mathcal{L}_{\tilde{S},\mathbf{e}(-\alpha)} \otimes_{\mathcal{O}_{\tilde{S}}} \Omega^q_{\tilde{S}}(\log D_{\tilde{S}})\big](-y)^{\lfloor n-\alpha\rfloor+q}, \tag{3}$$

where $\mathbf{e}(\alpha) := \exp(2\pi i\alpha)$, $\lfloor n-\alpha\rfloor$ denotes the integer part (see (27) below), and $td_{(1+y)*}$ is as in (13) below.

Here the sign $(-1)^{n-1}$ comes from the definition of spectrum, see (26) below. Note that the assertion for the Chern–Milnor class $M(X)$ corresponding to Theorem 1.1 (or deduced from it by specializing to $y=-1$) is essentially a corollary of [PaPr], and holds without assuming conditions (*a*), (*b*), see Sect. 3.5 below.

Condition (*a*) in Theorem 1.1 is satisfied if X is locally analytically trivial along each stratum S (e.g., if the intersection of X with transversal slices to any S has only isolated singularities of type A, D, E) or if the Hodge filtration F on any $\mathcal{H}^j_{S,\lambda} \otimes_{\mathbf{C}} \mathcal{O}_S$ is trivial (e.g., if every nonzero $\mathcal{H}^j_{S,\lambda}$ has rank 1). As for condition (*b*), we have the following:

Proposition 1.2. *Condition (b) is satisfied if the following two conditions hold:*

(*c*) *Every $S \in \mathcal{S}$ has a simply connected good compactification $\tilde{S}$.*
(*d*) *The local monodromy of $\mathcal{H}^j_{S,\lambda}$ around each irreducible component $D_{\tilde{S},i}$ of $D_{\tilde{S}} = \tilde{S} \setminus S$ is the multiplication by a constant number $c_{S,i,\lambda}$ which is independent of j.*

Conditions (a), (c), (d) are satisfied, for instance, in the case X is a projective hyperplane arrangement in $\mathbf{P}^n$ (see Proposition 1.4 below), or the projective compactification of the affine cone of a hypersurface in $\mathbf{P}^{n-1}$ which has only isolated singularities with semisimple Milnor monodromies. In these examples, the following conditions are satisfied for any $S \in \mathcal{S}$:

$$c_{S,i,\lambda} = \lambda^{-m'_{S,i}} \quad \text{with} \quad m'_{S,i} \in \mathbf{Z}, \tag{4}$$

$$X \cap Z_S = g_S^{-1}(0) \ \text{ with } \ g_S \ \text{ a homogeneous polynomial,} \tag{5}$$

where $Z_S \subset Y$ is an analytic transversal slice to S which intersects S transversally at a sufficiently general point of S. We assume that Z_S is sufficiently small, and has some coordinates to express g_S.

Set

$$\Lambda_S := \{\lambda \in \mathbf{C}^* \mid \mathcal{H}^j_{S,\lambda} \neq 0 \ (\exists j \in \mathbf{N})\}.$$

Let $G_S \subset \mathbf{C}^*$ be the subgroup generated by $\lambda \in \Lambda_S$. It corresponds to a finitely generated $\mathbf{Z}$-submodule of $\mathbf{Q}$ by $\alpha \mapsto \mathbf{e}(\alpha) = \exp(2\pi i\alpha)$, and is generated by $\mathbf{e}(1/m'_S)$ with $m'_S \in \mathbf{Z}_{>0}$. Note that the $m'_{S,i}$ are well-defined mod m'_S, and we sometimes assume

$$m'_{S,i}/m'_S \in [0,1). \tag{6}$$

However, it is not necessarily easy to give m'_S *explicitly* in general (even in the hyperplane arrangement case). If condition (5) is satisfied, then $\mathcal{H}^j_{S,\lambda} = 0$ unless $\lambda^{m_S} = 1$, and we get

$$m_S := \deg g_S \in \mathbf{Z}\, m'_S. \tag{7}$$

Here the equality $m_S = m'_S$ does not always hold (e.g., if $f = y_1^2 y_2^2$ with $m_S = 4$, $m'_S = 2$). We will assume (6) with m'_S replaced by m_S in case m'_S is not explicitly given.

Proposition 1.3. *(i) Assume conditions (c), (d) and (4), (6) hold. Then there is a rank* 1 *local system L'_S on S such that the eigenvalues of its local monodromies $c_{S,i,\lambda}$ satisfy (4) with $\lambda = \mathbf{e}(1/m'_S)$ and we have $L_{S,\lambda} \cong L'^{\otimes k}_S$ for any $\lambda = \mathbf{e}(k/m'_S) \in \Lambda_S$. Let $\mathcal{L}'_{\tilde{S}}$ be the Deligne extension of L'_S on $\tilde{S}$ such that the eigenvalues of the residues are contained in* $[0,1)$. *Then*

$$\mathcal{L}_{\tilde{S},\lambda} = \mathcal{L}'^{\otimes k}_{\tilde{S}} \otimes_{\mathcal{O}_{\tilde{S}}} \mathcal{O}_{\tilde{S}}\Big(\sum_i (\lceil k\, m'_{S,i}/m'_S\rceil - 1) D_{\tilde{S},i}\Big) \quad \textit{for } \lambda = \mathbf{e}(k/m'_S) \in \Lambda_S, \tag{8}$$

where $\lceil * \rceil$ *is as in (27) below.*

(ii) *With the above assumption, assume further that condition (5) holds and there is a rank* 1 *local system* L_S *on* S *such that the eigenvalues of its local monodromies* $c_{S,i,\lambda}$ *satisfy (4) with* $\lambda = \mathbf{e}(1/m_S)$. *Then the assertion of (i) together with (6) holds with* L'_S, $\mathcal{L}'_{\tilde{S}}$, m'_S *replaced, respectively, by* L_S, $\mathcal{L}_{\tilde{S}}$, m_S.

There is a certain similarity between (8) and [BuSa2, 1.4.3]. The difference between $\lceil * \rceil - 1$ in (8) and $\lfloor * \rfloor$ in loc. cit. comes from the difference between $j_!$ and $\mathbf{R}j_*$ if j denotes the open embedding $S \hookrightarrow \tilde{S}$. (It is also related to [BuSa1, Theorem 4.2] and [Sa2, 3.10.9]).

In the hyperplane arrangement case, an explicit formula for the $n_{f,S,\alpha}$ is given by [BuSa2] in the case X is reduced and $\mathrm{codim}_Y S \leqslant 3$, and we have the following (see Propositions 4.2 and 4.7 below):

Proposition 1.4. *If* X *is a projective hyperplane arrangement in* $\mathbf{P}^n$, *then conditions (a), (c), (d) and (4), (5) are all satisfied, and* L_S *exists so that (6) and (8) hold with* $\mathcal{L}'_{\tilde{S}}$, m'_S *replaced by* $\mathcal{L}_{\tilde{S}}$, m_S. *Moreover,* $\tilde{S}$, $\mathcal{L}_{\tilde{S}}$, $m'_{S,i}$ *are described explicitly, and the Hirzebruch–Milnor class is a combinatorial invariant of the hyperplane arrangement.*

The proof of the last assertion follows from an argument similar to [BuSa2], where the combinatorial property of the spectrum is shown by using the Hirzebruch–Riemann–Roch theorem together with [DCPr]. Here we use $\mathbf{H}_k(X) := \mathrm{CH}_k(X)_{\mathbf{Q}}$, since the structure of $\mathrm{CH}_k(\Sigma)_{\mathbf{Q}}$ is quite simple (see Proposition 4.6 below).

In Sect. 2 we review some basics of Hirzebruch characteristic classes and Steenbrink spectra of hypersurfaces. In Sect. 3 we give the proofs of Theorem 1.1 and Propositions 1.2 and 1.3. In Sect. 4 we treat the hyperplane arrangement case.

2 Preliminaries

In this section we review some basics of Hirzebruch characteristic classes and Steenbrink spectra of hypersurfaces.

2.1 Cohomology Hirzebruch Classes

In [Hi], Hirzebruch introduced the cohomology Hirzebruch characteristic class $T_y^*(TX)$ of the tangent bundle TX of a compact complex manifold X of dimension n. By using the formal Chern roots $\{\alpha_i\}$ for TX satisfying

$$\prod_{i=1}^{n}(1+\alpha_i t) = \sum_{j=0}^{n} c_j(TX)\, t^j,$$

it can be defined by

$$T_y^*(TX) := \prod_{i=1}^{n} Q_y(\alpha_i) \in \mathbf{H}^\bullet(X)[y]. \tag{9}$$

Here we use the formal power series in $\mathbf{Q}[y][[\alpha]]$

$$Q_y(\alpha) := \frac{\alpha(1+y)}{1-e^{-\alpha(1+y)}} - \alpha y, \quad \tilde{Q}_y(\alpha) := \frac{\alpha(1+ye^{-\alpha})}{1-e^{-\alpha}}, \quad R_y(\alpha) = \frac{e^{\alpha(1+y)}-1}{e^{\alpha(1+y)}+y}, \tag{10}$$

see [Hi, 10.2 and 11.1]. These have the relation

$$Q_y(\alpha) = (1+y)^{-1}\,\tilde{Q}_y(\alpha(1+y)) = \alpha/R_y(\alpha). \tag{11}$$

By specializing to $y=-1, 0, 1$, the power series $Q_y(\alpha)$ becomes, respectively,

$$1+\alpha, \quad \alpha/(1-e^{-\alpha}), \quad \alpha/\tanh\alpha,$$

and hence $T_y^*(TX)$ specializes to the Chern class $c^*(TX)$, the Todd class $td^*(TX)$, and the Thom–Hirzebruch L-class $L^*(TX)$, see [HiBeJu, Sect. 5.4].

2.2 *Homology Hirzebruch Classes*

The cohomology class $T_y^*(TX)$ is identified by Poincaré duality with the (Borel–Moore) homology class $T_y^*(TX) \cap [X]$, and this gives the definition of the homology Hirzebruch class $T_{y*}(X)$ in the smooth case. It is generalized to the singular case by [BrScYo]. Here we can use either the du Bois complex in [dB] or the bounded complex of mixed Hodge modules $\mathbf{Q}_{h,X}$ whose underlying $\mathbf{Q}$-complex is the constant sheaf $\mathbf{Q}_X$ in [Sa2].

Let $\mathrm{MHM}(X)$ be the abelian category of mixed Hodge modules on a complex algebraic variety X (see [Sa1, Sa2]). For $\mathcal{M}^\bullet \in D^b\mathrm{MHM}(X)$, its homology Hirzebruch characteristic class is defined by

$$\begin{aligned} &T_{y*}(\mathcal{M}^\bullet) := td_{(1+y)*}\big(\mathrm{DR}_y[\mathcal{M}^\bullet]\big) \in \mathbf{H}_\bullet(X)\left[y, \frac{1}{y(y+1)}\right] \quad \text{with} \\ &\mathrm{DR}_y[\mathcal{M}^\bullet] := \sum_{i,p} (-1)^i \left[\mathcal{H}^i \mathrm{Gr}_F^p \mathrm{DR}(\mathcal{M}^\bullet)\right](-y)^p \in K_0(X)[y, y^{-1}], \end{aligned} \tag{12}$$

where $\mathbf{H}_k(X) := H_{2k}^{\mathrm{BM}}(X, \mathbf{Q})$ or $\mathrm{CH}_k(X)_{\mathbf{Q}}$, and

$$td_{(1+y)*} : K_0(X)[y, y^{-1}] \to \mathbf{H}_\bullet(X)\left[y, \frac{1}{y(y+1)}\right] \tag{13}$$

is given by the scalar extension of the Todd class transformation

$$td_* : K_0(X) \to \mathbf{H}_\bullet(X),$$

(which is denoted by τ in [BaFuMa]) followed by the multiplication by $(1+y)^{-k}$ on the degree k part (see [BrScYo]). The last multiplication is closely related with the first equality of (11). By [Sch, Proposition 5.21], we have

$$T_{y*}(\mathcal{M}^\bullet) \in \mathbf{H}_\bullet(X)[y, y^{-1}].$$

The homology Hirzebruch characteristic class $T_{y*}(X)$ of a complex algebraic variety X is defined by applying the above definition to the case $\mathcal{M}^\bullet = \mathbf{Q}_{h,X}$ (see [BrScYo]), i.e.,

$$T_{y*}(X) := T_{y*}(\mathbf{Q}_{h,X}) = td_{(1+y)*}\mathrm{DR}_y[X] \in \mathbf{H}_\bullet(X)[y], \quad \text{with}$$
$$\mathrm{DR}_y[X] := \mathrm{DR}_y[\mathbf{Q}_{h,X}].$$

This coincides with the definition using the du Bois complex [dB]. It is known that $T_{y*}(X)$ belongs to $\mathbf{H}_\bullet(X)[y]$, see [BrScYo]. In case X is smooth, we have

$$\mathrm{DR}_y[X] = \Lambda_y[T^*X], \tag{14}$$

where we set for a vector bundle V on X

$$\Lambda_y[V] := \sum_{p\geqslant 0} [\Lambda^p V]\, y^p \in K^0(X)[y], \tag{15}$$

In fact, we have

$$\mathrm{DR}(\mathbf{Q}_{h,X}) = \mathrm{DR}(\mathcal{O}_X)[-n] = \Omega_X^\bullet \quad \text{with } \ n := \dim X, \tag{16}$$

where the Hodge filtration F^p on $\Omega_X^\bullet$ is defined by the truncation $\sigma_{\geqslant p}$ as in [De2]. (For the proof of the coincidence with the above definition of $T_{y*}(X)$ in the smooth case, we have to use the first equality of (11) and some calculation about Hirzebruch's power series $Q_y(\alpha)$ as in [HiBeJu, Sect. 5.4], or in the proof of [Yo1], Lemma 2.3.7, which is closely related with the generalized Hirzebruch–Riemann–Roch theorem as in [Hi, Theorem 21.3.1].)

2.3 Virtual Hirzebruch Classes

Hirzebruch [Hi] also introduced the virtual T_y-genus (or T_y-characteristic) which gives the T_y-genus of smooth complete intersections in smooth projective varieties.

Let X be any complete intersection in a smooth projective variety Y. We can define the virtual Hirzebruch characteristic class $T_{y*}^{\mathrm{vir}}(X)$ by

$$T_{y*}^{\mathrm{vir}}(X) := td_{(1+y)*}\mathrm{DR}_y^{\mathrm{vir}}[X] \in \mathbf{H}_\bullet(X)[y], \tag{17}$$

with $\mathrm{DR}_y^{\mathrm{vir}}[X]$ the image in $K_0(X)[[y]]$ of

$$\Lambda_y(T^*_{\mathrm{vir}}X) = \Lambda_y[T^*Y|_X]/\Lambda_y[N^*_{X/Y}] \in K^0(X)[[y]]. \tag{18}$$

Here $N^*_{X/Y}$ is defined by the locally free sheaf $\mathcal{I}_X/\mathcal{I}_X^2$ on X with $\mathcal{I}_X \subset \mathcal{O}_Y$ the ideal sheaf of the subvariety X of Y, and we set for a virtual vector bundle V on X

$$\Lambda_y V := \sum_{p\geqslant 0} \Lambda^p V\, y^p \in K^0(X)[[y]]. \tag{19}$$

Note that $\mathrm{DR}_y^{\mathrm{vir}}[X]$ belongs to $K_0(X)[y]$, see [MaSaSc1, Proposition 3.4]. (We denote by $K^0(X)$ and $K_0(X)$ the Grothendieck group of locally free sheaves of finite length and that of coherent sheaves, respectively.)

We have the equality $T_{y*}(X) = T_{y*}^{\mathrm{vir}}(X)$ if X is smooth. So the problem is how to describe their difference in the singular case, and this is given by the Hirzebruch–Milnor class as is explained in the introduction where only the hypersurface case is treated, see [MaSaSc1] for the complete intersection case. (For the degree-zero part, i.e., on the level of Hodge polynomials, see also [LiMa].)

2.4 Relation with Hirzebruch's Construction [Hi]

The image of the above virtual T_y-characteristic class of X by the trace morphism $\mathrm{Tr}_X : \mathbf{H}_\bullet(X) \to \mathbf{Q}$ coincides with the virtual T_y-genus of X constructed in [Hi, 11.2], where X is a (global) complete intersection of codimension r in a smooth complex projective variety Y with $i : X \hookrightarrow Y$ the natural inclusion. For this we have to recall the cohomological transformation T_y^* (as in [CaMaScSh] in the hypersurface case) applied to the virtual tangent bundle

$$T_{\mathrm{vir}}X := [TY|_X] - [N_{X/Y}] \in K^0(X).$$

This can be defined by

$$T_y^*(T_{\mathrm{vir}}X) := \left.\frac{\prod_i Q_y(\alpha_i)}{\prod_j Q_y(\beta_j)}\right|_X = \left.\frac{T_y^*(TY)}{\prod_j Q_y(\beta_j)}\right|_X. \tag{20}$$

Here the α_i are the formal Chern roots of TY, and the β_j ($j \in [1,r]$) are the cohomology classes of hypersurfaces of Y whose intersection is X.

By [MaSaSc1, Proposition 1.3.1] we have the equality

$$T_{y*}^{\mathrm{vir}}(X) = T_y^*(T_{\mathrm{vir}}X) \cap [X]. \tag{21}$$

This can be shown by using the first equality of (11) together with an argument similar to [Yo1, Lemma 2.3.7]. By the projection formula, we then get

$$i_*\big(T_{y*}^{\mathrm{vir}}(X)\big) = \frac{T_y^*(TY)}{\prod_j Q_y(\beta_j)} \cap i_*[X]. \tag{22}$$

We have moreover

$$i_*[X] = (\beta_1 \cup \cdots \cup \beta_r) \cap [Y]. \tag{23}$$

This follows, for instance, from the compatibility of the cycle class map $\mathrm{CH}^\bullet(Y) \to H^{2\bullet}(Y)$ with the multiplicative structures, see [Fu]. (Here X is defined scheme-theoretically by using a regular sequence, and we have $[X] = \sum_k m_k[X_k]$ with X_k the reduced irreducible components of X and m_k the multiplicities. The equality (23) is well known in the X smooth case.) Recall that we have by (11)

$$R_y(\beta_j) = \frac{\beta_j}{Q_y(\beta_j)}.$$

We thus get

$$i_*\big(T_{y*}^{\mathrm{vir}}(X)\big) = \left(\prod_j R_y(\beta_j) \cup T_y^*(TY)\right) \cap [Y]. \tag{24}$$

By applying the trace morphism $\mathrm{Tr}_{\mathrm{Y}} : H_\bullet(Y) \to \mathbf{Q}$, this implies the compatibility with Hirzebruch's construction [Hi, 11.2]

$$\mathrm{Tr}_X\big(T_{y*}^{\mathrm{vir}}(X)\big) = \int_Y \prod_j R_y(\beta_j) \cup T_y^*(TY). \tag{25}$$

Here $\int_Y : H^{2\dim Y}(Y) \to \mathbf{Q}$ denotes the canonical morphism (which is also called the trace morphism), and the right-hand side of (25) is equal to $T_y(\beta_1, \ldots, \beta_r)_Y$ in the notation of [Hi, 11.2] where Y and β_j are, respectively, denoted by M and v_j.

Note that the above argument is mostly useful for the degree-zero part of the (homology) Hirzebruch class unless the natural morphism $i_* : \mathbf{H}_\bullet(X) \to \mathbf{H}_\bullet(Y)$ is injective since the information may be lost in the other case.

2.5 Spectrum

Let f be a holomorphic function on a complex manifold Y of dimension n. Let $x \in X := f^{-1}(0) \subset Y$. We have the Steenbrink spectrum

$$\begin{aligned} \mathrm{Sp}(f,x) &= \sum_{\alpha\in\mathbf{Q}} n_{f,x,\alpha}\, t^{\alpha} \quad \text{with} \quad n_{f,x,\alpha} := \sum_j (-1)^{j-n+1} h^{p,j-p}_{f,x,\lambda}, \\ h^{p,j-p}_{f,x,\lambda} &:= \dim \mathrm{Gr}^p_F \tilde{H}^j(F_{f,x},\mathbf{C})_\lambda \quad \big(p = \lfloor n-\alpha \rfloor, \quad \lambda = \exp(-2\pi i\alpha)\big). \end{aligned} \tag{26}$$

Here F is the Hodge filtration of the canonical mixed Hodge structure on the reduced Milnor cohomology $\tilde{H}^j(F_{f,x},\mathbf{C})$ with $F_{f,x}$ the Milnor fiber of f around x, and $\tilde{H}^j(F_{f,x},\mathbf{C})_\lambda$ is the λ-eigenspace of the cohomology by the semisimple part T_s of the Milnor monodromy T, see [St1, St2]. Recall that

$$\lfloor\alpha\rfloor := \max\{\, i \in \mathbf{Z} \mid i \leqslant \alpha\}, \quad \lceil\alpha\rceil := \min\{\, i \in \mathbf{Z} \mid i \geqslant \alpha\}. \tag{27}$$

Let $i_x : \{x\} \hookrightarrow X$ denote the inclusion. Then we have an isomorphism of mixed Hodge structures

$$\tilde{H}^j(F_{f,x},\mathbf{Q}) = H^j i_x^* \varphi_f \mathbf{Q}_{h,Y}, \tag{28}$$

which is compatible with the action of the semisimple part T_s of the Milnor monodromy T. Here the category of mixed Hodge modules on a point is identified with the category of graded-polarizable mixed $\mathbf{Q}$-Hodge structures [De2], see [Sa2]. In fact, (28) is actually the definition of the mixed Hodge structure on the left-hand side.

3 Proofs of the Main Assertions

In this section we give the proofs of Theorem 1.1 and Propositions 1.2 and 1.3.

3.1 Proof of Proposition 1.2

Fix a base point $s_0 \in S$. Associated with the local system $\mathcal{H}^j_{S,\lambda}$, we have the monodromy representation

$$\rho^j_{S,\lambda} : \pi_1(S,s_0) \to \mathrm{Aut}(\mathcal{H}^j_{S,\lambda,s_0}). \tag{29}$$

Any $\gamma \in \pi_1(S,s_0)$ is represented by a piecewise linear path (using local coordinates). It is contractible inside $\tilde{S}$ by condition (c). We may assume that this contraction is also given by a piecewise linear one, and intersects $D_{\tilde{S}} = \tilde{S} \setminus S$ transversally at smooth points. (Here it may be better to use a sufficiently fine triangulation of $\tilde{S}$ compatible with $D_{\tilde{S}}$.) Then, using condition (*d*), we see that the image of γ by $\rho^j_{S,\lambda}$ is given by the multiplication by

$$\prod_i (c_{S,i,\lambda})^{a_i(\gamma)} \in \mathbf{C}^*, \tag{30}$$

since the multiplication by $c_{S,i,\lambda}$ belongs to the center of $\mathrm{Aut}(\mathcal{H}^j_{S,\lambda,s_0})$. Here $a_i(\gamma) \in \mathbf{Z}$ depends only on the contraction of γ, and is independent of j.

This argument implies that we have a monodromy representation

$$\rho_{S,\lambda} : \pi_1(S, s_0) \to \mathbf{C}^* \bigl(= \mathrm{Aut}(\mathbf{C})\bigr), \tag{31}$$

such that any nonzero $\rho^j_{S,\lambda}$ is isomorphic to a direct sum of copies of $\rho_{S,\lambda}$. Let $L_{S,\lambda}$ be the local system corresponding to $\rho_{S,\lambda}$. Then the assertion follows. This finishes the proof of Proposition 1.2.

3.2 *Proof of Proposition 1.3*

By condition (4) together with (30), the monodromy representation (31) is compatible with the product between the $\lambda \in \Lambda_S$. Note that Λ_S is stable by inverse. (In fact, the local systems $\mathcal{H}^j_S$ are defined over $\mathbf{Q}$ so that Λ_S is stable by complex conjugation, and the eigenvalues of the Milnor monodromies are roots of unity.) We then get the monodromy representation for any $\lambda \in G_S$

$$\rho_{S,\lambda} : \pi_1(S, s_0) \to \mathbf{C}^*, \tag{32}$$

in a compatible way with the product between the $\lambda \in G_S$. We define L'_S to be the rank 1 local system corresponding to $\rho_{S,\mathbf{e}(1/m'_S)}$.

By conditions (4) and (6) the eigenvalues of the residues of the connection of $\mathcal{L}'^{\otimes k}_{\tilde{S}}$ along $D_{\tilde{S},i}$ are given by

$$k m'_{S,i}/m'_S. \tag{33}$$

In fact, we have $m'_{S,i}/m'_S \in [0, 1)$ for $k = 1$ by condition (6). So (8) follows, and the assertion (i) is proved. The argument is similar for the assertion (ii). This finishes the proof of Proposition 1.3.

3.3 Proof of Theorem 1.1

Let $\mathcal{M}^j_{\tilde{S},\lambda}$ be the Deligne extension of the local system $\mathcal{H}^j_{S,\lambda}$ such that the eigenvalues of the residues of the connection are contained in $(0, 1]$ (see [De1]). By [MaSaSc1], Propositions 5.1.1 and 5.2.1, we have

$$
\begin{aligned}
&T_{y*}\big((i_{\Sigma\setminus X',\Sigma})_!\,\varphi_f\mathbf{Q}_{h,Y\setminus X'}\big)\\
&\quad= \sum_{S,j,p,q,\lambda} (-1)^{j+q}\,(\pi_{\tilde{S}})_* td_{(1+y)*}\big[\mathrm{Gr}^p_F\mathcal{M}^j_{\tilde{S},\lambda}\otimes_{\mathcal{O}_{\tilde{S}}}\Omega^q_{\tilde{S}}(\log D_{\tilde{S}})\big](-y)^{p+q},
\end{aligned}
\tag{34}
$$

By conditions (a) and (b) together with (26) and (28), we then get

$$
\begin{aligned}
&T_{y*}\big((i_{\Sigma\setminus X',\Sigma})_!\,\varphi_f\mathbf{Q}_{h,Y\setminus X'}\big)\\
&\quad= \sum_{S,j,p,q,\lambda} (-1)^{j+q}\,h^{p,j-p}_{f,S,\lambda}\,(\pi_{\tilde{S}})_* td_{(1+y)*}\big[\mathcal{L}_{\tilde{S},\lambda}\otimes_{\mathcal{O}_{\tilde{S}}}\Omega^q_{\tilde{S}}(\log D_{\tilde{S}})\big](-y)^{p+q},
\end{aligned}
\tag{35}
$$

with

$$
h^{p,j-p}_{f,S,\lambda} := h^{p,j-p}_{f,x,\lambda} \quad \text{for any } x\in S.
$$

In fact, condition (a) implies that the Hodge filtration F is defined on the level of local systems, and the graded pieces of the filtration F are still direct sums of rank 1 local systems as in condition (b). The assertion now follows from (35) and (26). This finishes the proof of Theorem 1.1.

3.4 Proof of [MaSaSc1, Proposition 4.1] in the Case $m > 1$

By the same argument as in [MaSaSc1], the assertion is reduced to the normal crossing case. Then it is enough to show the vanishing of the reduced cohomology of

$$
U_{\varepsilon,t} := \left\{ (y_1,\dots,y_n)\in\mathbf{C}^n \;\middle|\; \sum_{i=1}^{n-1}|y_i|^2<\varepsilon^2,\ |y_n|<\varepsilon,\ g=y_n^m\,t \right\}\ (0<|t|\ll\varepsilon\ll 1),
$$

by using the fundamental neighborhood system of $0\in\mathbf{C}^n$ given by

$$
U_\varepsilon := \left\{ (y_1,\dots,y_n)\in\mathbf{C}^n \;\middle|\; \sum_{i=1}^{n-1}|y_i|^2<\varepsilon^2,\ |y_n|<\varepsilon \right\}\ (0<\varepsilon\ll 1),
$$

where $g = \prod_{i=1}^{r} y_i^{m_i}$ with $r < n$. Consider

$$V_{\varepsilon,t} := \left\{ (y_1, \dots, y_{n-1}) \in \mathbf{C}^{n-1} \;\middle|\; \sum_{i=1}^{n-1} |y_i|^2 < \varepsilon^2,\ |g| < \varepsilon^m |t| \right\} \quad (0 < |t| \ll \varepsilon \ll 1).$$

It is contractible (see [Mi]), and $U_{\varepsilon,t}$ is a ramified covering of $V_{\varepsilon,t}$ which is ramified over the normal crossing divisor $V_{\varepsilon,t} \cap g^{-1}(0)$. Then $V_{\varepsilon,t}$ and $U_{\varepsilon,t}$ retract to $V_{\varepsilon,t} \cap g^{-1}(0)$, and the assertion follows.

3.5 *Formula for the Chern–Milnor Classes*

Let $M(X) \in H_\bullet(\Sigma)$ be the Chern–Milnor class $M(X)$ as in [PaPr] (which can be obtained by specializing $M_y(X)$ to $y = -1$). Then, without assuming conditions (a), (b), we have

$$\begin{aligned} M(X) &= \sum_{S \in \mathcal{S}} \tilde{\chi}(F_{f,S})\, c_*(1_S) \\ &= \sum_{S \in \mathcal{S}} \tilde{\chi}(F_{f,S})\, (\pi_{\tilde{S}})_* \big(c^* \big(\Omega^1_{\tilde{S}}(\log D_{\tilde{S}})^\vee \big) \cap [\tilde{S}] \big). \end{aligned} \tag{36}$$

Here $c_*(1_S) \in H_\bullet(\Sigma)$ is as in [Mac], and

$$\tilde{\chi}(F_{f,S}) := \chi(F_{f,S}) - 1,$$

with $F_{f,S}$ the Milnor fiber of f around a sufficiently general $x \in S$. The second equality of (36) follows from [Al, GoPa]. Note that the $\tilde{\chi}(F_{f,S})$ $(S \in \mathcal{S})$ give the constructible function associated with the vanishing cycle complex $\varphi_f \mathbf{Q}_{Y \setminus X'}$.

It is well known that if the restriction of X to a transversal slice to S at x is locally defined by a homogeneous polynomial g_S of degree m_S, then

$$\chi(F_{f,S}) = \chi\big(\mathbf{P}^{c_S - 1} \setminus g_S^{-1}(0)\big)\, m_S, \tag{37}$$

where $c_S := \operatorname{codim} S$.

In the hyperplane arrangement case, it is known (see [ScTeVa]) that

$$\chi\big(\mathbf{P}^{c_S - 1} \setminus g_S^{-1}(0)\big) = 0 \text{ if and only if } \bar{S} \text{ is not a dense edge,} \tag{38}$$

in the notation of (4.1) below.

4 Hyperplane Arrangement Case

In this section we treat the hyperplane arrangement case.

Notation 4.1. Assume X is a projective hyperplane arrangement in $Y = \mathbf{P}^n$, where $L = \mathcal{O}_{\mathbf{P}^n}$ and X' is a sufficiently general hyperplane. Let X_j be the irreducible components of X with multiplicities m_j ($j = 1, \dots, r$). Note that

$$m = \sum_j m_j,$$

and

$$X_j \subset \Sigma \quad \text{if} \quad m_j > 1.$$

We have a canonical stratification $\mathcal{S} = \mathcal{S}_{\Sigma'}$ of $\Sigma' := \Sigma \setminus X'$ such that

$$\bar{S} = \bigcap_{j \in I(S)} X_j, \quad S = \bar{S} \setminus \Bigl(\bigcup_{j \notin I(S)} X_j \cup X'\Bigr) \quad \text{with} \quad I(S) := \{j \mid X_j \supset S\}. \tag{39}$$

For the proof of Proposition 1.4, we have to consider also the canonical stratification $\mathcal{S}_X$ of X such that (39) holds by deleting X'. Here $\bar{S}$ is called an *edge* of the hyperplane arrangement X.

Let $C(X)$ denote the corresponding central hyperplane arrangement of $V := \mathbf{C}^{n+1}$ with irreducible components $C(X_j)$ and multiplicities m_j. Here $C(X_j)$ denotes the cone of X_j. Set

$$V^S := V/V_S \quad \text{with} \quad V_S := C(\bar{S}).$$

For each $S \in \mathcal{S}$, we have the quotient central hyperplane arrangement

$$C(X)^S \subset V^S$$

defined by the affine hyperplanes

$$C(X_j)^S := C(X_j)/V_S \subset V^S \quad \text{for } j \in I(S),$$

where the irreducible components $C(X_j)^S$ have the induced multiplicities m_j.

For each $S \in \mathcal{S}$, we also have the induced projective hyperplane arrangement

$$X_{\bar{S}} := \bigcup_{j \notin I(S)} X_j \cap \bar{S} \subset \bar{S},$$

such that $X_{\bar{S}} \setminus \Sigma$ has the induced stratification

$$\mathcal{S}_S := \{S' \in \mathcal{S} \mid S' \subset X_{\bar{S}}\}.$$

Here each $S' \in \mathcal{S}_S$ (especially for $\mathrm{codim}_{\bar{S}} S' = 1$) has the induced multiplicity

$$m_{S',S} := \sum_{j \in I(S') \setminus I(S)} m_j.$$

Let f^S be a homogeneous polynomial defining $C(X)^S \subset V^S$. This is essentially identified with g_S in (5), and

$$m_S = \deg f^S = \sum_{j \in I(S)} m_j \in \mathbf{Z}\, m'_S, \tag{40}$$

where m'_S is as in (6). Note that there is a shift of indices of the spectral numbers

$$n_{f,S,\alpha} = (-1)^{\dim S} n_{f^S,0,\beta} \quad \text{with} \;\; \beta = \alpha - \dim S, \tag{41}$$

and a formula for $n_{f^S,0,\beta}$ in the reduced case with $\dim V^S \leqslant 3$ can be found in [BuSa2].

As for the good compactification $\tilde{S}$ of $S \in \mathcal{S}$, it can be obtained by blowing-up $\bar{S} \subset \mathbf{P}^n$ along the edges $\bar{S}'$ of $X_{\bar{S}} \subset \bar{S}$ with $\mathrm{codim}_{\bar{S}} \bar{S}' \geqslant 2$ by decreasing induction on the codimension of the edges, where we restrict to the S' such that $X_{\bar{S}}$ is not a divisor with normal crossings on any neighborhood of S' in $\bar{S}$ as in [BuSa2]. (However, we do not restrict to the dense edges as in [ScTeVa], see Remark 4.4(i) below for the definition of dense edge.)

Let $\tilde{E}_{\bar{S}',\tilde{S}}$ be the proper transform of the exceptional divisor of the blow-up along $\bar{S}'$ in $\bar{S}$ if $\mathrm{codim}_{\bar{S}} \bar{S}' \geqslant 2$ (where we assume that $X_{\bar{S}} \subset \bar{S}$ is not a divisor with normal crossings on any neighborhood of S' as above). If $\mathrm{codim}_{\bar{S}} \bar{S}' = 1$, then $\tilde{E}_{\bar{S}',\tilde{S}}$ denotes the proper transform of $\bar{S}' \subset \bar{S}$. Let $\tilde{X}_{\infty,\tilde{S}}$ be the proper transform of $X_{\infty,\bar{S}} := X' \cap \bar{S} \subset \bar{S}$. These are the irreducible components $D_{\tilde{S},i}$ of $D_{\tilde{S}} \subset \tilde{S}$. The integers $m'_{S,i}$ in (4) for the components $\tilde{E}_{\bar{S}',\tilde{S}}$ and $\tilde{X}_{\infty,\tilde{S}}$ will be denoted, respectively, by

$$m'_{S',S}, \quad m'_{\infty,S}.$$

Proposition 4.2. *With the above notation, condition (4) holds with*

$$m'_{S',S}/m_S = \{m_{S',S}/m_S\}, \quad m'_{\infty,S}/m_S = \{-m/m_S\}, \tag{42}$$

where $\{\alpha\} := \alpha - \lfloor \alpha \rfloor$. *Moreover,* L_S *in Proposition 2 (ii) exists, and we have*

$$\mathcal{L}_{\tilde{S}} = \pi_{\tilde{S}}^{*}\mathcal{O}_{\bar{S}}\left(-\left\lceil \sum_{j\notin I(S)} m_j/m_S \right\rceil\right) \otimes_{\mathcal{O}_{\tilde{S}}} \mathcal{O}_{\tilde{S}}\left(\sum_{S'\subset S} \lfloor m_{S',S}/m_S \rfloor \tilde{E}_{\bar{S}',\tilde{S}}\right), \tag{43}$$

where $\mathcal{O}_{\bar{S}}(k)$ *denotes the pull-back of* $\mathcal{O}_{\mathbf{P}^n}(k)$ *by* $\bar{S} \hookrightarrow \mathbf{P}^n$.

Proof. By using the blowing-ups along the $\bar{S}'$ in $\mathbf{P}^n$, the assertion (42) is reduced to Lemma 4.3 below. The integrable connection corresponding to the local system L_S can be constructed easily on $S \subset \bar{S} \setminus X' \cong \mathbf{C}^{\dim S}$, see, e.g., [EsScVi].

For the proof of (43), it is then enough to show

$$\mathcal{L}_{\tilde{S}}^{\otimes m_S} = \pi_{\tilde{S}}^{*}\mathcal{O}_{\bar{S}}\left(-m_S\left\lceil \sum_{j\notin I(S)} m_j/m_S \right\rceil\right) \otimes_{\mathcal{O}_{\tilde{S}}} \mathcal{O}_{\tilde{S}}\left(\sum_{S'\subset S} m_S\lfloor m_{S',S}/m_S \rfloor\, \tilde{E}_{\bar{S}',\tilde{S}}\right). \tag{44}$$

(Indeed, the simply connectedness implies that $\mathrm{CH}^1(\tilde{S})$ is torsion-free, since it implies that $H^1(\tilde{S},\mathbf{Z}) = 0$ and $H^2(\tilde{S},\mathbf{Z})$ is torsion-free.) The left-hand side of (44) is a line bundle with a logarithmic connection such that the eigenvalues of the residues along $\tilde{E}_{\bar{S}',\tilde{S}}$ and $\tilde{X}_{\infty,\tilde{S}}$ are, respectively, $m'_{S',S}$ and $m'_{\infty,S}$ [since $m'_{S',S}/m_S,\ m'_{\infty,S}/m_S \in [0,1)$ by (42)]. So we get

$$\mathcal{L}_{\tilde{S}}^{\otimes m_S} = \mathcal{O}_{\tilde{S}}\left(-\sum_{S'\subset S} m'_{S',S}\tilde{E}_{\bar{S}',\tilde{S}} - m'_{\infty,S}\tilde{X}_{\infty,\tilde{S}}\right).$$

On the other hand, setting $X_{j,\bar{S}} := X_j \cap \bar{S}$, we have

$$\begin{aligned}\pi_{\tilde{S}}^{*}\mathcal{O}_{\bar{S}}\left(-m_S\left\lceil \sum_{j\notin I(S)} m_j/m_S \right\rceil\right) &= \pi_{\tilde{S}}^{*}\mathcal{O}_{\bar{S}}\left(-\sum_{j\notin I(S)} m_j X_{j,\bar{S}} - m'_{\infty,S}X_{\infty,\bar{S}}\right)\\ &= \mathcal{O}_{\tilde{S}}\left(-\sum_{S'\subset S} m_{S',S}\tilde{E}_{\bar{S}',\tilde{S}} - m'_{\infty,S}\tilde{X}_{\infty,\tilde{S}}\right).\end{aligned}$$

Here the first isomorphism follows from the second equality of (42) (which implies that $m'_{\infty,S} \in [0, m_S)$), since we have by the definition of m and m_S

$$\sum_{j\notin I(S)} m_j = m \mod m_S.$$

The second isomorphism is obtained by calculating the total transform of the divisor. So the assertion follows from (42) which implies that

$$m'_{S',S} = m_{S',S} - m_S\lfloor m_{S',S}/m_S \rfloor.$$

Lemma 4.3. *Let g be a holomorphic function on a complex manifold Y. Set $X := Y\times\mathbf{C}^*$, and $f := gz^a$ where z is the coordinate of $\mathbf{C}$ and $a \in \mathbf{Z}$. Then the monodromy of the local system $\mathcal{H}^j\psi_f\mathbf{Q}_X|_{\{y\}\times\mathbf{C}^*}$ for $y \in Y$ is given by T^{-a}, where T is the Milnor monodromy.*

Proof. Since $f^{-1}(0)$ is analytic-locally trivial along $\{y\}\times\mathbf{C}^*$, we can calculate $\mathcal{H}^j(\psi_f\mathbf{Q}_X)_{(y,z)}$ by the cohomology of

$$\left\{y' \in Y \;\middle|\; ||y'|| < \varepsilon,\ g(y') = z^{-a}t\right\} \quad (0 < |t| \ll \varepsilon \ll 1).$$

Here $||y'||$ is defined by taking local coordinates of Y around y, and we may assume $|z| = 1$ for the calculation of the monodromy. Then the assertion is clear. This finishes the proofs of Lemma 4.3 and Proposition 4.2.

Remarks 4.4. (i) An edge $\bar{S}$ of a hyperplane arrangement X is called *dense* (see [ScTeVa]) if its associated quotient central hyperplane arrangement $C(X)^S$ is indecomposable. Note that a central hyperplane arrangement is called indecomposable if it is not a union of hyperplane arrangements coming from $\mathbf{C}^{n_1}$ and $\mathbf{C}^{n_1}$ via the projections

$$\mathbf{C}^n \to \mathbf{C}^{n_1}, \quad \mathbf{C}^n \to \mathbf{C}^{n_2},$$

where $n = n_1 + n_2$ and $n_1, n_2 > 0$.

(ii) If Y is a simply connected smooth variety and $Z \subset Y$ is a closed subvariety of codimension at least two, then $Y \setminus Z$ is simply connected. (Indeed, a contraction of a path has real dimension 2, and can be modified so that it does not intersect Z.)

This implies that if Y is a simply connected smooth variety and $Y' \to Y$ is a proper birational morphism from a smooth variety, then Y' is also simply connected. (Indeed, if a smooth variety has a dense Zariski-open subvariety which is simply connected, then it is also simply connected. This follows from the fact that a path has real dimension 1, and may be modified so that it is contained in the open subvariety.)

(iii) The image of the cycle map is contained in

$$\mathrm{Gr}^W_{-2k}H_{2k}(\Sigma, \mathbf{Q}) \subset H_{2k}(\Sigma, \mathbf{Q}),$$

and so is the image of td_*. Moreover, the structures of $\mathrm{Gr}^W_{-2k}H_{2k}(\Sigma, \mathbf{Q})$ and $\mathrm{CH}_k(X)_{\mathbf{Q}}$ in the projective hyperplane arrangement case are quite simple as below.

Proposition 4.5. *Let X be a projective hyperplane arrangement in $\mathbf{P}^n$ with X_j $(j \in [1, r])$ the irreducible components. Let W be the weight filtration of the canonical mixed Hodge structure on $H_k(X, \mathbf{Q})$ (which is the dual of $H^k(X, \mathbf{Q})$). Then we have*

$$\mathrm{Gr}^W_{-k} H_k(X, \mathbf{Q}) = \begin{cases} \bigoplus_{1 \leqslant j \leqslant r} \mathbf{Q}[X_j] & \textit{if } k = 2n-2, \\ \mathbf{Q}(k/2) & \textit{if } k \in 2\mathbf{Z} \cap [0, 2n-4], \\ 0 & \textit{otherwise}, \end{cases} \tag{45}$$

where $[X_j]$ *denotes the class of* X_j (*and* $\mathbf{Q}[X_j]$ *is not a polynomial algebra*). *Moreover we have the canonical isomorphisms*

$$\mathrm{Gr}^W_{-k} H_k(X, \mathbf{Q}) \xrightarrow{\sim} H_k(\mathbf{P}^n, \mathbf{Q})(k/2) \quad (0 \leqslant k < 2n-2). \tag{46}$$

Proof. We have a long exact sequence

$$\to H^{\mathrm{BM}}_{k+1}(\mathbf{P}^n \setminus X, \mathbf{Q}) \to H_k(X, \mathbf{Q}) \to H_k(\mathbf{P}^n, \mathbf{Q}) \to,$$

together with

$$H^{\mathrm{BM}}_{k+1}(\mathbf{P}^n \setminus X, \mathbf{Q}) = H^{2n-k-1}(\mathbf{P}^n \setminus X, \mathbf{Q})(n).$$

Here $H^p(\mathbf{P}^n \setminus X, \mathbf{Q})$ vanishes for $p > n$, and has type (p, p) for $p \leqslant n$ by [Bri]. (In fact, an integral logarithmic p-form has type (p, p) by [De4, Theorem 8.2.4 (i)], since the latter implies that it induces a morphism of mixed Hodge structures from $\mathbf{Q}(-p)$ to the pth cohomology group.)

Setting $p = 2n - k - 1$, we see that $H^{\mathrm{BM}}_{k+1}(\mathbf{P}^n \setminus X, \mathbf{Q}) = 0$ if $k + 1 < n$, and it has type

$$(n-k-1, n-k-1), \ \text{ if } \ k+1 \geqslant n.$$

Here we have

$$2n - 2k - 2 > -k \quad \text{for} \quad k \in [0, 2n-3].$$

So we get (46) and also (45) except for $k = 2n - 2$. In the last case, the above argument shows that $H_{2n-2}(X, \mathbf{Q})$ has type $(1-n, 1-n)$, and has dimension r by using [Bri]. So the assertion follows.

Proposition 4.6. *With the notation of 4.1 and Proposition 4.5 above, set*

$$\Sigma_1 := \bigcup_{X_j \subset \Sigma} X_j, \quad \Sigma_2 := \overline{\Sigma \setminus \Sigma_1}, \quad \mathcal{S}^{(i)} := \{S \in \mathcal{S} \mid \mathrm{codim}_Y S = i\},$$

$$\mathcal{S}_a := \{S \in \mathcal{S} \mid S \in \Sigma_a\}, \quad \mathcal{S}_a^{(i)} := \mathcal{S}_a \cap \mathcal{S}^{(i)} \quad (a = 1, 2).$$

Then we have

$$\mathrm{CH}_k(X)_{\mathbf{Q}} = \begin{cases} \bigoplus_{1\leqslant j\leqslant r} \mathbf{Q}[X_j] & \textit{if } k = n-1, \\ \mathbf{Q} & \textit{if } k \in [0, n-2]. \end{cases} \tag{47}$$

$$\mathrm{CH}_k(\Sigma)_{\mathbf{Q}} = \begin{cases} \bigoplus_{S\in\mathcal{S}_1^{(1)}} \mathbf{Q}[\bar{S}] & \textit{if } k = n-1, \\ \bigoplus_{S\in\mathcal{S}_2^{(2)}} \mathbf{Q}[\bar{S}] \oplus \mathbf{Q} & \textit{if } k = n-2 \textit{ and } \Sigma_1 \neq \emptyset, \\ \bigoplus_{S\in\mathcal{S}_2^{(2)}} \mathbf{Q}[\bar{S}] & \textit{if } k = n-2 \textit{ and } \Sigma_1 = \emptyset, \\ \mathbf{Q} & \textit{if } k \in [0, n-3]. \end{cases} \tag{48}$$

Moreover, for $S \in \mathcal{S}^{(i)}$ with $i \geqslant 3$, we have the canonical isomorphisms

$$\mathrm{CH}_k(\bar{S})_{\mathbf{Q}} \xrightarrow{\sim} \mathrm{CH}_k(\Sigma)_{\mathbf{Q}} \xrightarrow{\sim} \mathrm{CH}_k(\mathbf{P}^n)_{\mathbf{Q}} \quad (0 \leqslant k \leqslant \dim \bar{S} \leqslant n-3), \tag{49}$$

$$\mathrm{CH}_{n-2}(\bar{S})_{\mathbf{Q}} \xrightarrow{\sim} \mathrm{CH}_{n-2}(\Sigma_1)_{\mathbf{Q}} \xrightarrow{\sim} \mathrm{CH}_k(\mathbf{P}^n)_{\mathbf{Q}} \quad \textit{if} \quad S \in \mathcal{S}_1^{(2)}, \tag{50}$$

and $\mathbf{Q}$ in (48) for $k = n-2$ and $\Sigma_1 \neq \emptyset$ is given by the image of $\mathrm{CH}_{n-2}(\Sigma_1)_{\mathbf{Q}}$ in (50).

Proof. Since $\dim \Sigma_1 = n-1$ and $\dim \Sigma_2 = n-2$, the assertions easily follow from the well-known facts that we have for $\bar{S}' \subset \bar{S} \subset \mathbf{P}^n$

$$\mathrm{CH}_k(\bar{S}) = \mathbf{Q} \quad \text{for } k \in [0, \dim \bar{S}], \tag{51}$$

$$\mathrm{CH}_k(\bar{S}') \xrightarrow{\sim} \mathrm{CH}_k(\bar{S})) \xrightarrow{\sim} \mathrm{CH}_k(\mathbf{P}^n) \text{ for } k \in [0, \dim \bar{S}']. \tag{52}$$

For instance, (52) implies that the image of $[\bar{S}']$ in $\mathrm{CH}_{n-2}(\Sigma_1)_{\mathbf{Q}}$ is independent of $\bar{S}' \in \mathcal{S}_1^{(2)}$ (by applying it to the $\bar{S} \in \mathcal{S}_1^{(1)}$). So (50) follows. The proofs of the other assertions are similar. This finishes the proof of Proposition 4.6.

Proposition 4.7. *The Hirzebruch–Milnor class $M_y(X)$ of a hyperplane arrangement X in $\mathbf{P}^n$ is a combinatorial invariant, where $\mathbf{H}_\bullet(\Sigma) = \mathrm{CH}_\bullet(\Sigma)_{\mathbf{Q}}$ and Proposition 4.6 is used.*

Proof. Let $\mathcal{E}_{\tilde{S},\alpha,q}$ be a vector bundle on $\tilde{S}$ defined by

$$\mathcal{E}_{\tilde{S},\alpha,q} := \mathcal{L}_{\tilde{S},\mathbf{e}(-\alpha)} \otimes_{\mathcal{O}_{\tilde{S}}} \Omega^q_{\tilde{S}}(\log D_{\tilde{S}}).$$

By Theorem 1.1, Proposition 4.6 and [BuSa2] (which implies that the $n_{f,S,\alpha}$ are combinatorial invariants) together with the definition of $td_{(1+y)*}$ in (12), it is enough to show that the following is a combinatorial invariant for any S, α, q :

$$(\pi_{\tilde{S},\bar{S}})_* td_*(\mathcal{E}_{\tilde{S},\alpha,q}) \in \mathbf{H}_\bullet(\bar{S}) = \mathbf{Q}^{\dim \bar{S}+1}. \tag{53}$$

Here $\pi_{\tilde{S},\bar{S}} : \tilde{S} \to \bar{S}$ is the canonical morphism, and (51) is used for the last isomorphism. Since $\bar{S}$ is projective space, we may assume for the proof of (53)

$$\mathbf{H}_\bullet(\bar{S}) = H_{2\bullet}(\bar{S}, \mathbf{Q}), \quad \mathbf{H}^\bullet(\bar{S}) = H^{2\bullet}(\bar{S}, \mathbf{Q}),$$

and similarly for $\tilde{S}$. By [BaFuMa] we have

$$td_*(\mathcal{E}_{\tilde{S},\alpha,q}) = \big(ch(\mathcal{E}_{\tilde{S},\alpha,q}) \cup td^*(T\tilde{S})\big) \cap [\tilde{S}] \quad \text{in} \quad \mathbf{H}_\bullet(\tilde{S}).$$

This means that $td_*(\mathcal{E}_{\tilde{S},\alpha,q})$ is identified by Poincaré duality with

$$ch(\mathcal{E}_{\tilde{S},\alpha,q}) \cup td^*(T\tilde{S}) \in \mathbf{H}^\bullet(\tilde{S}).$$

Moreover, the pushforward by $\pi_{\tilde{S},\bar{S}}$ is calculated by the top degree part of

$$ch(\mathcal{E}_{\tilde{S},\alpha,q}) \cup td^*(T\tilde{S}) \cup \pi^*_{\tilde{S},\bar{S}} e^k,$$

where e^k is the canonical generator of $\mathbf{H}^k(\bar{S})$ ($k \in [0, \dim \bar{S}]$). These can be calculated by the method in [BuSa2], Sect. 5 using the combinatorial description of the cohomology ring as in [DCPr]. Here $[\tilde{E}_{\bar{S}',\tilde{S}}]$, $[\pi^*_{\tilde{S},\bar{S}} e] \in \mathbf{H}^1(\tilde{S})$ correspond to e_V ($V \neq 0$) and $-e_0$ in [BuSa2, 5.3]. Moreover $c^*(\Omega^q_{\tilde{S}}(\log D_{\tilde{S}})$ and $td^*(T\tilde{S})$ are combinatorially expressed [BuSa2, 5.4] by using certain universal polynomials. Combining these with Propositions 1.3 (ii) and 4.2, it follows that (53) is a combinatorial invariant. Here it is not necessary to use the full result of [DCPr], since we need only the fact that the multiple-intersection numbers of the e_V are independent of the position of the irreducible components, and are determined combinatorially (together with certain relations among the e_V). This finishes the proof of Proposition 4.7.

Acknowledgements The first named author is partially supported by NSF-1304999. The second named author is partially supported by Kakenhi 24540039. The third named author is supported by the SFB 878 "groups, geometry and actions."

References

[Al] P. Aluffi, Differential forms with logarithmic poles and Chern-Schwartz-MacPherson classes of singular varieties. C. R. Acad. Sci. Paris Sér. I Math. **329**, 619–624 (1999)

[BaFuMa] P. Baum, W. Fulton, R. MacPherson, Riemann-Roch for singular varieties. Inst. Hautes Etudes Sci. Publ. Math. **45**, 101–145 (1975)

[Bri] E. Brieskorn, Sur les groupes de tresses, in *Séminaire Bourbaki (1971/1972), Exp. 401*. Lecture Notes in Mathematics, vol. 317 (Springer, Berlin, 1973), pp. 21–44

[BrScYo] J.-P. Brasselet, J. Schürmann, S. Yokura, Hirzebruch classes and motivic Chern classes of singular spaces. J. Topol. Anal. **2**, 1–55 (2010)

[BuSa1] N. Budur, M. Saito, Multiplier ideals, *V*-filtration, and spectrum. J. Algebra Geom. **14**, 269–282 (2005)

[BuSa2] N. Budur, M. Saito, Jumping coefficients and spectrum of a hyperplane arrangement. Math. Ann. **347**, 545–579 (2010)

[CaMaScSh] S.E. Cappell, L. Maxim, J. Schürmann, J.L. Shaneson, Characteristic classes of complex hypersurfaces. Adv. Math. **225**, 2616–2647 (2010)

[dB] Ph. du Bois, Complexe de de Rham filtré d'une variété singulière. Bull. Soc. Math. Fr. **109**, 41–81 (1981)

[DCPr] C. De Concini, C. Procesi, Wonderful models of subspace arrangements. Sel. Math. N. Ser. **1**, 459–494 (1995)

[De1] P. Deligne, *Equations Différentielles à Points Singuliers Réguliers*. Lecture Notes in Mathematics, vol. 163 (Springer, Berlin, 1970)

[De2] P. Deligne, Théorie de Hodge, II. Inst. Hautes Etudes Sci. Publ. Math. **40**, 5–57 (1971)

[De3] P. Deligne, Le formalisme des cycles évanescents, in *SGA7 XIII and XIV*. Lecture Notes in Mathematics, vol. 340 (Springer, Berlin, 1973), pp. 82–115, 116–164.

[De4] P. Deligne, Théorie de Hodge, III. Inst. Hautes Etudes Sci. Publ. Math. **44**, 5–77 (1974)

[DiMaSaTo] A. Dimca, Ph. Maisonobe, M. Saito, T. Torrelli, Multiplier ideals, *V*-filtrations and transversal sections. Math. Ann. **336**, 901–924 (2006)

[EsScVi] H. Esnault, V. Schechtman, E. Viehweg, Cohomology of local systems on the complement of hyperplanes. Invent. Math. **109**, 557–561 (1992)

[Fu] F. Fulton, *Intersection Theory* (Springer, Berlin, 1984)

[GoPa] M. Goresky, W. Pardon, Chern classes of automorphic vector bundles. Inv. Math. **147**, 561–612 (2002)

[Hi] F. Hirzebruch, *Topological Methods in Algebraic Geometry* (Springer, Berlin, 1966)

[HiBeJu] F. Hirzebruch, T. Berger, R. Jung, *Manifolds and Modular Forms*. Aspects of Mathematics, vol. E20 (Friedrich Vieweg & Sohn, Braunschweig, 1992)

[LiMa] A. Libgober, L. Maxim, Hodge polynomials of singular hypersurfaces. Michigan Math. J. **60**, 661–673 (2011)

[Mac] R.D. MacPherson, Chern classes for singular algebraic varieties. Ann. Math. (2) **100**, 423–432 (1974)

[MaSaSc1] L. Maxim, M. Saito, J. Schürmann, Hirzebruch-Milnor classes of complete intersections. Adv. Math. **241**, 220–245 (2013)

[MaSaSc2] L. Maxim, M. Saito, J. Schürmann, Spectral Hirzebruch-Milnor classes of singular hypersurfaces. Preprint arXiv:1606.02218

[Mi] J. Milnor, *Singular Points of Complex Hypersurfaces*. Annals of Mathematics Studies, vol. 61 (Princeton University Press, Princeton, NJ, 1968)

[PaPr] A. Parusiński, P. Pragacz, Characteristic classes of hypersurfaces and characteristic cycles. J. Alg. Geom. **10**, 63–79 (2001)

[Sa1] M. Saito, Modules de Hodge polarisables. Publ. RIMS Kyoto Univ. **24**, 849–995 (1988)

[Sa2] M. Saito, Mixed hodge modules. Publ. RIMS Kyoto Univ. **26**, 221–333 (1990)

[Sa3] M. Saito, Thom-Sebastiani theorem for Hodge modules. Preprint 2010

[ScTeVa] V. Schechtman, H. Terao, A. Varchenko, Local systems over complements of hyperplanes and the Kac-Kazhdan conditions for singular vectors. J. Pure Appl. Algebra **100**, 93–102 (1995)

[Sch] J. Schürmann, Characteristic classes of mixed hodge modules, in *Topology of Stratified Spaces*. MSRI Publications, vol. 58 (Cambridge University Press, Cambridge, 2011), pp. 419–471.

[St1] J.H.M. Steenbrink, Mixed hodge structure on the vanishing cohomology, in *Real and Complex Singularities*. Proceedings of Ninth Nordic Summer School, Oslo, 1976 (Sijthoff and Noordhoff, Alphen aan den Rijn, 1977), pp. 525–563

[St2] J.H.M. Steenbrink, The spectrum of hypersurface singularities. Astérisque **179–180**, 163–184 (1989)

[Yo1] S. Yokura, A generalized Grothendieck-Riemann-Roch theorem for Hirzebruch's χ_y-characteristic and T_y-characteristic. Publ. RIMS Kyoto Univ. **30**, 603–610 (1994)

[Yo2] S. Yokura, On characteristic classes of complete intersections, in *Algebraic Geometry: Hirzebruch 70 (Warsaw, 1998)*. Contemporary Mathematics, vol. 241 (American Mathematical Society, Providence, RI, 1999), pp. 349–369

On Lusztig's q-Analogues of All Weight Multiplicities of a Representation

Dmitri I. Panyushev

2010 *Mathematics Subject Classification.* 17B10, 17B20, 20G10

1 Introduction

The ground field $\Bbbk$ is algebraically closed and of characteristic zero. Let G be a connected semisimple algebraic group, and T a maximal torus inside a Borel subgroup B. Write $\mathfrak{g}, \mathfrak{t}$, and $\mathfrak{b}$ for their Lie algebras. If $\mathbb{V}$ is a finite-dimensional rational G-module, then $\mathbb{V} = \oplus_{\mu\in\mathfrak{t}^*}\mathbb{V}^\mu$ is the weight decomposition with respect to T (or $\mathfrak{t}$). If $\mathbb{V} = \mathbb{V}_\lambda$ is a simple G-module with highest weight λ, then $m_\lambda^\mu = \dim(\mathbb{V}_\lambda^\mu)$. In this article, we present some results on Lusztig's q-analogues $\mathfrak{M}_\lambda^\mu(q)$ of weight multiplicities m_λ^μ. The polynomial $\mathfrak{M}_\lambda^\mu(q)$ is defined algebraically as an alternating sum over the Weyl group, through the q-analogue of Kostant's partition function. Initially, Lusztig introduced q-analogues only for dominant weights μ [Lu83, (9.4)]. However, this constraint is unnecessary and $\mathfrak{M}_\lambda^\mu(q)$ is a non-trivial polynomial for any μ such that $\lambda - \mu$ is a linear combination of positive roots with nonnegative coefficients; in particular, for all weights of $\mathbb{V}_\lambda$. A relationship with certain Kazhdan–Lusztig polynomials [Ka82] implies that $\mathfrak{M}_\lambda^\mu(q)$ has nonnegative coefficients whenever μ is dominant. For instance, if $\mathbb{V}_\lambda$ has the zero weight, with $m_\lambda^0 = n$, then $\mathfrak{M}_\lambda^0(q) = \sum_{i=1}^n q^{m_i(\lambda)}$ and $m_1(\lambda), \dots, m_n(\lambda)$ are the *generalised exponents* of $\mathbb{V}_\lambda$. These numbers were first considered by

D.I. Panyushev (✉)
Institute for Information Transmission Problems of the R.A.S., Bol'shoi Karetnyi per. 19, 127994 Moscow, Russia

Independent University of Moscow, Bol'shoi Vlasevskii per. 11, 119002 Moscow, Russia
e-mail: panyushev@iitp.ru

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_10

Kostant [Ko63, n. 5] in connection with the graded G-module structure of the ring $\Bbbk[\mathcal{N}]$, where $\mathcal{N} \subset \mathfrak{g}$ is the nilpotent cone. The interpretation of Kostant's generalised exponents via polynomials $\mathfrak{M}_\lambda^0(q)$ is due to Hesselink [He80] and D. Peterson (unpublished).

In Sect. 2, we gather basic properties of the polynomials $\mathfrak{M}_\lambda^\mu(q)$ and recall their relationship to cohomology of line bundles on $G \times_B \mathfrak{u}$. We emphasise the role of results of Broer on the nonnegativity of coefficients of $\mathfrak{M}_\lambda^\mu(q)$ [Br93] and the induction lemma for computing $\mathfrak{M}_\lambda^\mu(q)$ [br94]. Using Broer's results allows us to quickly recover some known results on coefficients of degenerate Cherednik kernel that appear in work of Bazlov et al. [Ba01, Ion04, V06]. We also prove that $\mathfrak{M}_\lambda^\mu(q+1)$ is a polynomial in q with nonnegative coefficients.

In Sect. 3, $\mathfrak{g}$ is assumed to be simple, and then θ is the highest root. We determine Lusztig's q-analogues for all roots of $\mathfrak{g} = \mathbb{V}_\theta$. Furthermore, if $\mathfrak{g}$ has two root lengths, then the short dominant root θ_s determines a representation that is called *little adjoint*, and we also compute q-analogues for all weights of $\mathbb{V}_{\theta_s}$. Then we obtain a formula for the weighted sum $\sum_\mu m_\theta^\mu \mathfrak{M}_\theta^\mu(q)$, which implies that it depends only on $\mathfrak{M}_\theta^0(q)$ and the Coxeter number of $\mathfrak{g}$. A similar result is valid for $\mathbb{V}_{\theta_s}$.

In Sect. 4, we prove that, for any simple G-modules $\mathbb{V}_\lambda$ and $\mathbb{V}_\gamma$, the sum $\sum_\mu m_\gamma^\mu \mathfrak{M}_\lambda^\mu(q)$ is equal to the q-analogue of the zero weight multiplicity for the (reducible) G-module $\mathbb{V}_\lambda \otimes \mathbb{V}_\gamma^*$ (Theorem 4.2). Therefore, $\sum_\mu m_\gamma^\mu \mathfrak{M}_\lambda^\mu(q) = \sum_\mu m_\lambda^\mu \mathfrak{M}_\gamma^\mu(q)$ and this also provides another formula for the $\mathbb{Z}[q]$-valued symmetric bilinear form on the character ring of $\mathfrak{g}$ that was introduced by R. Gupta (Brylinski) in [G87-2]. As a by-product, we obtain that such a weighted sum is always a polynomial with nonnegative coefficients. Comparing two formulae for $\sum_\mu m_{\theta_s}^\mu \mathfrak{M}_{\theta_s}^\mu(q)$ yields a curious identity involving the Poincaré polynomial for W_{θ_s}, the Weyl group stabiliser of θ_s, and $\mathfrak{M}_{\theta_s}^0(q)$ (Corollary 4.6). We hope that there are other interesting results pertaining to q-analogues of all weights of a representation.

If $\mathfrak{g}$ is simple and η_i is the number of positive roots of height i, then the partition formed by the exponents of $\mathfrak{g}$ is dual (conjugate) to the partition formed by the η_i's, see [Ko59, Ion04, V06]. Section 5 contains a geometric explanation and generalisation to this result. Let $e \in \mathfrak{g}$ be a principal nilpotent element. We prove that if $\dim \mathbb{V}_\lambda^e = \dim \mathbb{V}_\lambda^\mathfrak{t}$, then the "positive" weights of $\mathbb{V}_\lambda$ exhibit the similar phenomenon relative to the generalised exponents of $\mathbb{V}_\lambda$.

1.1 Main Notation

Throughout, G is a connected semisimple algebraic group with Lie $G = \mathfrak{g}$. We fix a Borel subgroup B and a maximal torus $T \subset B$, and consider the corresponding triangular decomposition $\mathfrak{g} = \mathfrak{u} \oplus \mathfrak{t} \oplus \mathfrak{u}^-$, where Lie $B = \mathfrak{u} \oplus \mathfrak{t}$. Then

- Δ is the *root system* of $(\mathfrak{g}, \mathfrak{t})$, Δ^+ is the set of positive roots corresponding to $\mathfrak{u}$, $\Pi = \{\alpha_1, \dots, \alpha_r\}$ is the set of simple roots in Δ^+, and $\rho = \frac{1}{2}\sum_{\mu\in\Delta^+}\mu$;
- $\mathfrak{X}$ is the lattice of integral weights of T and $\mathfrak{t}_\mathbb{Q}^*$ is the $\mathbb{Q}$-vector subspace of $\mathfrak{t}^*$ generated by $\mathfrak{X}$, $Q = \oplus_{i=1}^r \mathbb{Z}\alpha_i \subset \mathfrak{X}$ is the *root lattice*, and Q_+ is the monoid

generated by $\alpha_1, \dots, \alpha_r$. If $\gamma = \sum_{i=1}^{r} c_i\alpha_i \in Q_+$, then $\mathsf{ht}(\gamma) = \sum_{i=1}^{r} c_i$ is the *height* of γ.
– $\mathfrak{X}_+$ is the monoid of dominant weights and $\varphi_i \in \mathfrak{X}_+$ is the fundamental weight corresponding to $\alpha_i \in \Pi$;
– W is the *Weyl group* of $(\mathfrak{g}, \mathfrak{t})$ and $(\ ,\ )$ is a W-invariant positive-definite inner product on $\mathfrak{t}^*_{\mathbb{Q}}$. As usual, $\mu^\vee = 2\mu/(\mu, \mu)$ is the coroot for $\mu \in \Delta$.
– If $\lambda \in \mathfrak{X}_+$, then $\mathbb{V}_\lambda$ is the simple G-module with highest weight λ, $\mathbb{V}^*_\lambda$ is its dual, and $\lambda^* \in \mathfrak{X}_+$ is defined by $\mathbb{V}_{\lambda^*} = \mathbb{V}^*_\lambda$.

For $\alpha \in \Pi$, we let s_α denote the corresponding simple reflection in W. If $\alpha = \alpha_i$, then we also write $s_i = s_{\alpha_i}$. The *length function* on W with respect to $s_1, \dots, s_r$ is denoted by ℓ.

2 Generalities on q-Analogues of Weight Multiplicities

If $\lambda \in \mathfrak{X}_+$, then $\mathbb{V}^\mu_\lambda$ is the μ-weight space of $\mathbb{V}_\lambda$, $m^\mu_\lambda = \dim \mathbb{V}^\mu_\lambda$, and $\chi_\lambda = \mathsf{ch}(\mathbb{V}_\lambda) = \sum_\mu m^\mu_\lambda e^\mu \in \mathbb{Z}[\mathfrak{X}]$ is the *character* of $\mathbb{V}_\lambda$. Let $\varepsilon(w) = (-1)^{\ell(w)}$ be the sign of $w \in W$. By Weyl's character formula, $\mathsf{ch}(\mathbb{V}_\lambda) = \dfrac{\sum_{w\in W} \varepsilon(w) e^{w(\lambda+\rho)}}{e^\rho \prod_{\gamma\in\Delta^+}(1-e^{-\gamma})}$. For $\mu, \gamma \in \mathfrak{X}$, we write $\mu \preccurlyeq \gamma$, if $\gamma - \mu \in Q_+$.

Define functions $\mathcal{P}_q(\mu)$ by the equation

$$\frac{1}{\prod_{\alpha\in\Delta_+}(1-qe^\alpha)} =: \sum_{\mu\in Q_+} \mathcal{P}_q(\mu)e^\mu .$$

Then $\mathcal{P}_q(\mu)$ is a polynomial in q with $\deg \mathcal{P}_q(\mu) = \mathsf{ht}(\mu)$ and $\mu \mapsto \mathcal{P}(\mu) := \mathcal{P}_q(\mu)|_{q=1}$ is the usual Kostant's partition function. For $\lambda, \mu \in \mathfrak{X}_+$, Lusztig [Lu83, (9.4)] (see also [Ka82, (1.2)]) introduced a fundamental q-analogue of weight multiplicities m^μ_λ:

$$\mathfrak{M}^\mu_\lambda(q) = \sum_{w\in W} \varepsilon(w)\mathcal{P}_q(w(\lambda+\rho) - (\mu+\rho)). \tag{1}$$

For series $\mathbf{A}_r$, these are the classical *Kostka–Foulkes polynomials*. Therefore, this name is sometimes used in the general situation. It is also known that $\mathfrak{M}^\mu_\lambda(q)$ are related to certain Kazhdan–Lusztig polynomials associated with the corresponding affine Weyl group [Lu83], [Ka82, Theorem 1.8]. However, one needn't restrict oneself with only dominant weights μ, and the polynomials $\mathfrak{M}^\mu_\lambda(q)$ can be considered for arbitrary $\mu \in \mathfrak{X}$. It is easily seen that

- $\mathfrak{M}^\mu_\lambda(q) \equiv 0$ unless $\lambda \succcurlyeq \mu$;
- if $\lambda \succcurlyeq \mu$, then $\mathfrak{M}^\mu_\lambda(q)$ is a monic polynomial and $\deg \mathfrak{M}^\mu_\lambda(q) = \mathsf{ht}(\lambda - \mu)$; therefore, $\mathfrak{M}^\lambda_\lambda(q) \equiv 1$;
- $\mathfrak{M}^\mu_\lambda(1) = m^\mu_\lambda$.

In particular, if $\mu \preccurlyeq \lambda$, but μ is not a weight of $\mathbb{V}_\lambda$, then $\mathfrak{M}_\lambda^\mu(1) = 0$ and therefore $\mathfrak{M}_\lambda^\mu(q)$ has negative coefficients. If μ is dominant, then the relationship with Kazhdan–Lusztig polynomials implies that $\mathfrak{M}_\lambda^\mu(q)$ has nonnegative coefficients. The most general result on nonnegativity of the coefficients of $\mathfrak{M}_\lambda^\mu(q)$, whose proof exploits the cohomological interpretation, is due to Broer [Br93], see Theorem 2.2 below.

2.1 A Relationship to Cohomology of Line Bundles

Let $\mathcal{Z}$ be the cotangent bundle of G/B, i.e., $\mathcal{Z} = G \times_B \mathfrak{u}$. Recall that the corresponding collapsing $\mathcal{Z} \to G\mathfrak{u} =: \mathcal{N} \subset \mathfrak{g}$ is birational and $H^0(\mathcal{Z}, \mathcal{O}_\mathcal{Z}) = \Bbbk[\mathcal{N}]$ [He76]. Here $\mathcal{N}$ is the cone of nilpotent elements of $\mathfrak{g}$. For $\mu \in \mathfrak{X}$, let $\Bbbk_\mu$ denote the corresponding one-dimensional B-module. We consider line bundles on $\mathcal{Z}$ induced from homogeneous line bundles on G/B, i.e., line bundles of the form

$$G \times_B (\mathfrak{u} \oplus \Bbbk_\mu) \to G \times_B \mathfrak{u} = \mathcal{Z}.$$

The (invertible) sheaf of sections of this bundle is denoted by $\mathcal{L}_\mathcal{Z}(\Bbbk_\mu)$. More generally, if N is a rational B-module, then

$$G \times_B (\mathfrak{u} \oplus N) \to G \times_B \mathfrak{u} = \mathcal{Z}$$

is a vector bundle on $\mathcal{Z}$ of rank $\dim N$ and the corresponding sheaf of sections (locally free $\mathcal{O}_\mathcal{Z}$-module) is $\mathcal{L}_\mathcal{Z}(N)$. If $\mathcal{E}$ is a locally free $\mathcal{O}_\mathcal{Z}$-module, then $\mathcal{E}^\star$ is its dual. For instance, $\mathcal{L}_\mathcal{Z}(N)^\star = \mathcal{L}_\mathcal{Z}(N^*)$, where N^* is the dual B-module.

The cohomology groups of $\mathcal{L}_\mathcal{Z}(N)$ have a natural structure of a graded G-module by

$$H^i(G \times_B \mathfrak{u}, \mathcal{L}_{G\times_B \mathfrak{u}}(N)) \simeq \bigoplus_{j=0}^{\infty} H^i(G/B, \mathcal{L}_{G/B}(\mathcal{S}^j\mathfrak{u}^* \otimes N)),$$

where $\mathcal{S}^j\mathfrak{u}^*$ is the j-th symmetric power of the dual of $\mathfrak{u}$. Set $H^i(\mu) := H^i(\mathcal{Z}, \mathcal{L}_\mathcal{Z}(\mu)^\star)$. It is a graded G-module with

$$(H^i(\mu))_j = H^i(G/B, \mathcal{L}_{G/B}(\mathcal{S}^j\mathfrak{u} \otimes \Bbbk_\mu)^\star).$$

As $\dim(H^i(\mu))_j < \infty$, the graded character of $H^i(\mu)$ is well defined:

$$\mathsf{ch}_q(H^i(\mu)) = \sum_j \sum_{\lambda \in \mathfrak{X}_+} \dim \mathsf{Hom}_G\big(\mathbb{V}_\lambda, (H^i(\mu))_j\big)\chi_\lambda q^j \in \mathbb{Z}[\mathfrak{X}][[q]].$$

The reader is referred to work of Broer and Brylinski for more details [Br93, br94, B89].

Theorem 2.1 ([B89, Lemma 6.1]). *For any* $\mu \in \mathfrak{X}$, *we have*

$$\sum_i (-1)^i \mathsf{chch}_q(H^i(\mu)) = \sum_{\lambda \in \mathfrak{X}_+} \mathfrak{M}_\lambda^\mu(q)\chi_\lambda^*.$$

A more general version of this relation, where $\mathfrak{n} \subset \mathfrak{g}$ is replaced with a B-stable subspace of an arbitrary G-module $\mathbb{V}_\lambda$, appears in [P10, Theorem 3.8].

For $\mu = 0$, we have $\mathcal{L}_\mathcal{Z}(0) = \mathcal{O}_\mathcal{Z}$ and $H^i(\mathcal{Z}, \mathcal{O}_\mathcal{Z}) = 0$ for $i > 0$ [He76]. Therefore, the sum $\sum_{\lambda \in \mathfrak{X}_+} \mathfrak{M}_\lambda^0(q)\chi_\lambda^*$ represents the graded character of $H^0(\mathcal{Z}, \mathcal{O}_\mathcal{Z}) \simeq \Bbbk[\mathcal{N}]$ [He80].

For $\mu \in \mathfrak{X}$, we write μ^+ for the unique element in $W\mu \cap \mathfrak{X}_+$.

Theorem 2.2 (Broer's Criterion [Br93, Br97]). *The following conditions are equivalent for* $\mu \in \mathfrak{X}$*:*

(1) $\mathfrak{M}_\lambda^\mu(q)$ *has nonnegative coefficients for all* $\lambda \in \mathfrak{X}_+$*;*
(2) *if* $\mu \preccurlyeq \gamma \preccurlyeq \mu^+$ *and* $\gamma \in \mathfrak{X}_+$*, then* $\gamma = \mu^+$*;*
(3) $(\mu, \nu^\vee) \geqslant -1$ *for all* $\nu \in \Delta^+$.

The equivalence of (1) and (2) is proved in [Br93, Theorem 2.4]; the underlying reason is that, for such μ, higher cohomology of $\mathcal{L}_\mathcal{Z}(\mu)^\star$ vanishes. The equivalence of (2) and (3) appears in [Br97, Proposition 2(iii)].

Remark 2.3. The required equivalence of (2) and (3) is correctly proved by Broer, but we have noticed some other assertions of Proposition 2 in [Br97] are false. Namely, in part (iii) Broer claims the equivalence of certain conditions (a)–(c), where (a) and (b) are just our conditions (2) and (3). But condition (c) must be excluded from that list. Moreover, part (ii) in [Br97, Proposition 2] is also false. In both cases, a counterexample is given, e.g., by $\beta = -\varphi_1$ for $\mathfrak{g} = \mathfrak{sl}_{r+1}$, $r \geqslant 2$. This β satisfies Broer's conditions (a) and (b), but not (c); and part (ii) also fails for β. More generally, if $\kappa \in \mathfrak{X}_+$ is minuscule, then $\beta = -\kappa$ provides a counterexample to Broer's assertions.

Recall that $m_\lambda^0 \neq 0$ if and only if $\lambda \in \mathfrak{X}_+ \cap Q$. Then $\mathfrak{M}_\lambda^0(q) = \sum_{j=1}^n q^{m_j(\lambda)}$ $(n = m_\lambda^0)$ is a polynomial with nonnegative coefficients and the integers $m_1(\lambda), \dots, m_n(\lambda)$ are called the *generalised exponents* of $\mathbb{V}_\lambda$. If $\mathfrak{g}$ is simple and $\mathbb{V}_\lambda = \mathfrak{g}$, then they coincide with the usual exponents of $\mathfrak{g}$ (= of W) [He80].

2.2 Broer's Induction Lemma and Degenerate Cherednik Kernel

The following fundamental result of Broer is a powerful tool for computing q-analogues of weight multiplicities. Unfortunately, it did not attract the attention it deserves. Perhaps the reason is that Broer formulates it as a relation in "the Grothendieck group of finitely generated graded $\Bbbk[\mathcal{N}]$-modules with a compatible G-module structure". However, extracting the coefficients of χ_λ^*, one obtains the following down-to-earth description:

Theorem 2.4 (Induction Lemma, cf. [br94, Proposition 3.15]). *Let* $\lambda \in \mathfrak{X}_+$. *If* $\gamma \in \mathfrak{X}$ *and* $(\gamma, \alpha^\vee) = -n < 0$ *for some* $\alpha \in \Pi$ *(hence* $s_\alpha(\gamma) = \gamma + n\alpha$*), then*

$$\mathfrak{M}_\lambda^\gamma(q) + \mathfrak{M}_\lambda^{s_\alpha(\gamma)-\alpha}(q) = q(\mathfrak{M}_\lambda^{\gamma+\alpha}(q) + \mathfrak{M}_\lambda^{s_\alpha(\gamma)}(q)). \tag{2}$$

In particular, for $n = 1$, this formula contains only γ and $\gamma + \alpha$ and one merely obtains $\mathfrak{M}_\lambda^\gamma(q) = q\,\mathfrak{M}_\lambda^{\gamma+\alpha}(q)$. Broer's proof of the Induction Lemma exploits the cohomological interpretation of Lusztig's q-analogues discussed above, and includes the passage from G/B to G/P_α, where P_α is the minimal parabolic subgroup corresponding to α.

Actually, the name "Induction Lemma" is assigned in [br94] to a certain preparatory result. But, we feel that it is more appropriate to associate such a name with Broer's Proposition 3.15.

It is observed in [G87-1, 5.1] that Lusztig's q-analogues $\mathfrak{M}_\lambda^\mu(q)$ satisfy the identity

$$\sum_{\mu:\,\mu\preccurlyeq\lambda} \mathfrak{M}_\lambda^\mu(q)e^\mu = \frac{\sum_{w\in W}\varepsilon(w)e^{w(\lambda+\rho)}}{e^\rho\prod_{\gamma\in\Delta^+}(1-qe^{-\gamma})} = \chi_\lambda\cdot\prod_{\gamma\in\Delta^+}\frac{(1-e^{-\gamma})}{(1-qe^{-\gamma})} = \chi_\lambda\xi_q\,. \tag{3}$$

Here $\xi_q = \prod_{\gamma\in\Delta^+}\dfrac{1-e^{-\gamma}}{1-qe^{-\gamma}}$ is the *degenerate Cherednik kernel*, and for $\lambda = 0$ one obtains

$$\xi_q = \sum_{\mu\in Q_+} \mathfrak{M}_0^{-\mu}(q)e^{-\mu}. \tag{4}$$

Thus, the coefficients of ξ_q are certain Lusztig's q-analogues. As an application of the Induction Lemma, we easily recover some known results on coefficients of ξ_q, cf. Bazlov [Ba01, Theorem 3], Ion [Ion04, Eq. (5.35)], and Viswanath [V06, Proposition 1]. We also write $[e^{-\mu}](\xi_q)$ for the coefficient of $e^{-\mu}$ in ξ_q.

Proposition 2.5. *If* $\mu \in \Delta^+$, *then* $[e^{-\mu}](\xi_q) = \mathfrak{M}_0^{-\mu}(q) = q^{\mathsf{ht}(\mu)} - q^{\mathsf{ht}(\mu)-1}$.

Proof. We argue by induction on $\mathsf{ht}(\mu)$.

1) *Base*: if $\mu \in \Delta^+$ is simple, then it is easily seen that $[e^{-\mu}](\xi_q) = q - 1$.
2) *Step*: Suppose that $\mathsf{ht}(\mu) \geqslant 2$ and the assertion holds for all $\gamma \in \Delta^+$ with $\mathsf{ht}(\gamma) < \mathsf{ht}(\mu)$. Take any $\alpha \in \Pi$ such that $(\mu, \alpha^\vee) = n > 0$. Then $s_\alpha(\mu) = \mu - n\alpha \in \Delta^+$ and applying (2) with $\gamma = -\mu$ we obtain

$$\mathfrak{M}_0^{-\mu}(q) + \mathfrak{M}_\lambda^{-\mu+(n-1)\alpha}(q) = q(\mathfrak{M}_\lambda^{-\mu+\alpha}(q) + \mathfrak{M}_\lambda^{-\mu+n\alpha}(q)).$$

Since $\mu - \alpha$ and $\mu - (n-1)\alpha$ are also positive roots, of smaller height, using the induction assumption yields the desired expression for $\mathfrak{M}_0^{-\mu}(q)$. □

Remark 2.6. Bazlov and Ion work with the usual (2-parameter) Cherednik kernel, and then specialise their formulae to one-parameter case. They use the general theory of Macdonald polynomials, whereas Viswanath provides a direct elementary approach to computing coefficients of ξ_q. One can notice that Viswanath's note [V06] contains implicitly an inductive formula for the coefficients of ξ_q. His argument essentially proves that if $\beta \in Q_+$ and $s_i(\beta) = \beta - k\alpha_i$ ($k > 0$), then

$$\mathfrak{M}_0^{-\beta}(q) = (q-1)\sum_{j=1}^{k-1} \mathfrak{M}_0^{-\beta+j\alpha_i}(q) + q\cdot\mathfrak{M}_0^{-s_i(\beta)}(q). \tag{5}$$

Actually, one needn't assume here that $s_i(\beta) \in Q_+$. If some of $\beta - j\alpha_i$ do not belong to Q_+, then the corresponding q-analogues are replaced by zero. It is a simple exercise to deduce (5) from (2) with $\lambda = 0$, and vice versa. [Left to the reader.]

Substituting (4) in the equality $\sum_{\mu:\, \mu\preccurlyeq\lambda} \mathfrak{M}_\lambda^\mu(q)e^\mu = \chi_\lambda\xi_q$, we obtain

$$\mathfrak{M}_\lambda^\mu(q) = \sum_{\gamma:\, \gamma\succcurlyeq\mu} m_\lambda^\gamma \mathfrak{M}_0^{\mu-\gamma}(q), \tag{6}$$

so that all q-analogues for $\mathbb{V}_\lambda$ can (theoretically) be computed once we know enough coefficients of ξ_q and the usual weight multiplicities. But even for the adjoint representation, this approach requires more than merely the knowledge of $\mathfrak{M}_0^{-\nu}(q)$ for $\nu \in \Delta^+$. For, $\gamma - \mu$ need not be a root in the above formula. However, Eq. (6) has a curious consequence.

Lemma 2.7. *The polynomials $\mathfrak{M}_\lambda^\mu(q+1)$ have nonnegative coefficients for all μ. If μ is a weight of $\mathbb{V}_\lambda$ and $\mu \neq \lambda$, then $\mathfrak{M}_\lambda^\mu(0) = 0$.*

Proof. 1) By the very definition of ξ_q, we have $\xi_q = \prod_{\gamma\in\Delta^+} \bigl(1+\sum_{n\geqslant 0} q^n(q-1)e^{(n+1)\gamma}\bigr)$.
Whence all polynomials $\mathfrak{M}_0^\nu(q+1)$, the coefficients of ξ_{q+1} have nonnegative coefficients. Using Eq. (6), we carry it over to arbitrary $\lambda \in \mathfrak{X}_+$.

2) By Weyl's denominator formula, $\xi_q|_{q=0} = \sum_{w\in W} \varepsilon(w)e^{w\rho-\rho}$. Therefore, $\mathfrak{M}_0^\nu(0) = \varepsilon(w)$ if $\nu = w\rho - \rho$, and is zero otherwise. Hence $\mathfrak{M}_\lambda^\mu(0) = \sum_{w\in W} \varepsilon(w)m_\lambda^{\mu+\rho-w\rho}$. For a weight μ of $\mathbb{V}_\lambda$, the latter equals $\delta_{\lambda\mu}$ by Klimyk's formula, see, e.g., [S90, § 3.8, Proposition C].
[One can also refer directly to Eq. (1).] □

3 All q-Analogues for the Adjoint and Little Adjoint Representations

In this section, $\mathfrak{g}$ is simple, θ is the highest root, and θ_s is the short dominant root in Δ^+. Here we compute q-analogues for all weight multiplicities of the adjoint and little adjoint representations of $\mathfrak{g}$ and show that their sum depends essentially only on the q-analogue of the zero weight multiplicity and the Coxeter number of $\mathfrak{g}$.

Let $m_i = m_i(\theta)$, $i = 1, \dots, r$, be the *exponents* of (the adjoint representation of) $\mathfrak{g}$ and h the *Coxeter number* of $\mathfrak{g}$. We assume that $m_1 \leqslant m_2 \leqslant \dots \leqslant m_r$, hence $m_1 = 1$ and $m_r = h - 1 = \mathsf{ht}(\theta)$. In the simply laced case, all roots are assumed to be short. That is, the argument referring to long roots has to be omitted if $\mathfrak{g}$ is of type A-D-E.

Theorem 3.1. *For any $\mu \in \Delta \cup \{0\}$, the polynomial $\mathfrak{M}^{\mu}_{\theta}(q)$ depends only on $\mathsf{ht}(\theta - \mu)$, i.e., on $\mathsf{ht}(\mu)$. More precisely,*

(i) $\mathfrak{M}^{0}_{\theta}(q) = q^{m_1} + \dots + q^{m_r}$;
(ii) *If $\mu \in \Delta^+$, then $\mathfrak{M}^{\mu}_{\theta}(q) = q^{\mathsf{ht}(\theta-\mu)} = q^{h-1-\mathsf{ht}(\mu)}$;*
(iii) *if $\alpha \in \Pi$, then $\mathfrak{M}^{-\alpha}_{\theta}(q) = (q-1)\mathfrak{M}^{0}_{\theta}(q) + q^{h-1}$;*
(iv) *If $\mu \in \Delta^+$, then $\mathfrak{M}^{-\mu}_{\theta}(q) = q^{\mathsf{ht}(\mu)-1} \cdot \mathfrak{M}^{-\alpha}_{\theta}(q)$.*

Proof. (i) This is well known and goes back to Hesselink [He80] and Peterson. See also [Ion04, Theorem 5.5] and [V06, p. 2].

(ii) If $\mu \in \Delta^+$ is short, then $(\mu, \gamma^\vee) \geqslant -1$ for all $\gamma \in \Delta^+$ and therefore $\mathfrak{M}^{\mu}_{\theta}(q)$ has nonnegative coefficients by Broer's criterion (Theorem 2.2). Since $\deg \mathfrak{M}^{\mu}_{\theta}(q) = \mathsf{ht}(\theta - \mu)$ and $m^{\mu}_{\theta} = 1$, one has the only possibility for $\mathfrak{M}^{\mu}_{\theta}(q)$.

If Δ has two root lengths and $\mu \in \Delta^+$ is long, then we argue by induction in $\mathsf{ht}(\theta - \mu)$. For $\mu = \theta$, one has $\mathfrak{M}^{\theta}_{\theta}(q) = 1$. To perform the induction step, assume that $\mathfrak{M}^{\mu}_{\theta}(q) = q^{\mathsf{ht}(\theta-\mu)}$ for some μ and $\mu \notin \Pi$. Then there is $\alpha \in \Pi$ such that $(\alpha, \mu) > 0$ and hence $s_\alpha(\mu) \in \Delta^+$ and $\mathsf{ht}(s_\alpha(\mu)) < \mathsf{ht}(\mu)$. Here $\mu = s_\alpha(\mu) + n\alpha$ with $n \in \{1, 2, 3\}$, and by the Induction Lemma (Theorem 2.4) applied to $\gamma = s_\alpha(\mu)$ we have

$$\mathfrak{M}^{s_\alpha(\mu)}_{\theta}(q) + \mathfrak{M}^{\mu-\alpha}_{\theta}(q) = q(\mathfrak{M}^{s_\alpha(\mu)+\alpha}_{\theta}(q) + \mathfrak{M}^{\mu}_{\theta}(q)).$$

For $n = 1$, we immediately obtain that $\mathfrak{M}^{s_\alpha(\mu)}_{\theta}(q) = q\mathfrak{M}^{\mu}_{\theta}(q) = q^{\mathsf{ht}(\theta - s_\alpha(\mu))}$. For $n = 2$ or 3, we get the same conclusion using the fact that the roots $s_\alpha(\mu) + \alpha$ and $\mu - \alpha$ are short (and hence the corresponding q-analogues are already known).

(iii) Passing from $\alpha \in \Pi$ to $-\alpha$ (crossing over 0) is also accomplished via the use of the Induction Lemma. Since $s_\alpha(-\alpha) = -\alpha + 2\alpha$, we have

$$\mathfrak{M}^{-\alpha}_{\theta}(q) + \mathfrak{M}^{0}_{\theta}(q) = q(\mathfrak{M}^{0}_{\theta}(q) + \mathfrak{M}^{\alpha}_{\theta}(q)),$$

and it is already proved in part (ii) that $\mathfrak{M}^{\alpha}_{\theta}(q) = q^{\mathsf{ht}(\theta-\alpha)} = q^{h-2}$.

(iv) Going down from $-\alpha$ ($\alpha \in \Pi$), we again use the Induction Lemma. First, we prove the assertion for all negative short roots using the fact that if $\mu \in \Delta^+$ is short and $\mu \neq \theta_s$, then there is $\alpha \in \Pi$ such that $(\mu, \alpha^\vee) = -1$ and hence $s_\alpha(-\mu) = -\mu - \alpha$. Afterwards, we prove the assertion for the long roots, as it was done in part (ii).

□

Remark 3.2. The simplest formula for $\mathfrak{M}_\theta^{-\alpha}(q)$, $\alpha \in \Pi$, occurs if $\mathfrak{g} = \mathfrak{sl}_{r+1}$, where there are only three summands. Namely, $\mathfrak{M}_\theta^{-\alpha}(q) = q^{r+1} + q^r - q$. But for $\mathfrak{g} = \mathfrak{sp}_{2r}$ or $\mathfrak{so}_{2r+1}$, $r \geqslant 2$, we obtain $\mathfrak{M}_\theta^{-\alpha}(q) = q^2 + q^4 + \cdots + q^{2r} - (q + q^3 + \cdots + q^{2r-3})$.

The notation $\mu \dashv \mathbb{V}_\lambda$ means that μ is a weight of $\mathbb{V}_\lambda$.

Theorem 3.3. *We have* $\sum_{\mu \dashv \mathfrak{g}} \mathfrak{M}_\theta^\mu(q) = \mathfrak{M}_\theta^0(q)(\mathfrak{M}_\theta^0(q) - r + 1) + \frac{\mathfrak{M}_\theta^0(q)}{q} \cdot \frac{1-q^h}{1-q}$ *or, equivalently,*

$$\sum_{\mu \dashv \mathfrak{g}} m_\theta^\mu \mathfrak{M}_\theta^\mu(q) = \mathfrak{M}_\theta^0(q)^2 + \frac{\mathfrak{M}_\theta^0(q)}{q} \cdot \frac{1-q^h}{1-q}. \tag{7}$$

Proof. Since $m_\theta^\mu = 1$ for $\mu \in \Delta$ and $m_\theta^0 = r$, both formulae are equivalent. In fact, we compute separately the sums

$$\mathcal{S}_+ = \sum_{\mu \in \Delta^+} \mathfrak{M}_\theta^\mu(q) \quad \text{and} \quad \mathcal{S}_- = \sum_{\mu \in \Delta^-} \mathfrak{M}_\theta^\mu(q).$$

Recall that the partition $(m_r, \ldots, m_1)$ is dual to the partition $(n_1, n_2, \ldots)$, where $n_i = \#\{\gamma \in \Delta^+ \mid \mathsf{ht}(\gamma) = i\}$ [Ko59]. Therefore, Δ^+ can be partitioned into the strings of roots of lengths $m_1, m_2, \ldots, m_r$ such that the i-th string contains the roots of height $1, 2, \ldots, m_i$. Then, by Theorem 3.1(ii), the sum over the i-th string equals

$$q^{m_r-1} + q^{m_r-2} + \cdots + q^{m_r-m_i} = \frac{q^{m_r-m_i} - q^{m_r}}{1-q}.$$

Since $m_i + m_{r-i+1} = m_r + 1 = h$, the total sum over Δ^+ can be written as

$$\mathcal{S}_+ = \sum_{i=1}^{r} \frac{q^{m_i-1} - q^{h-1}}{1-q} = \sum_{i=1}^{r} \frac{q^{m_i} - q^h}{q(1-q)} = \frac{\mathfrak{M}_\theta^0(q) - rq^h}{q(1-q)}.$$

Likewise, using the corresponding strings of negative roots, one proves that

$$\mathcal{S}_- = ((q-1)\mathfrak{M}_\theta^0(q) + q^{h-1}) \cdot \frac{r - \mathfrak{M}_\theta^0(q)}{1-q}.$$

It then remains to simplify the sums $\mathcal{S}_+ + \mathfrak{M}_\theta^0(q) + \mathcal{S}_-$ and $\mathcal{S}_+ + r\mathfrak{M}_\theta^0(q) + \mathcal{S}_-$. □

Similar results are valid for the little adjoint representation of G. Let Δ_s denote the set of all short roots, hence $\{\theta_s\} = \Delta_s \cap \mathfrak{X}_+$. Set $\Pi_s = \Pi \cap \Delta_s$ and $l = \#(\Pi_s)$. Recall that the set of weights of $\mathbb{V}_{\theta_s}$ is $\Delta_s \cup \{0\}$, $m_{\theta_s}^0 = l$, and $m_{\theta_s}^\mu = 1$ for $\mu \in \Delta_s$.

The following observation is a particular case of [Ion04, Theorem 5.5], and we provide a proof for the reader's convenience.

Lemma 3.4. *Let* $(n_{1s}, n_{2s}, \dots)$ *be the partition of* $\#(\Delta_s^+)$ *with* $n_{i,s} = \#\{\gamma \in \Delta_s^+ \mid \mathsf{ht}(\gamma) = i\}$*, in particular,* $\eta_{1,s} = l$*. If* $(e_1, e_2, \dots, e_l)$ *is the dual partition, then* $\mathfrak{M}^0_{\theta_s}(q) = q^{e_1} + \dots + q^{e_l}$*.*

Proof. By Proposition 2.5 and (6), we have

$$\mathfrak{M}^0_{\theta_s}(q) = \sum_{\gamma \in Q_+} m^{\gamma}_{\theta_s} \mathfrak{M}^{-\gamma}_0(q) = l + \sum_{\mu \in \Delta_s^+} (q^{\mathsf{ht}(\mu)} - q^{\mathsf{ht}(\mu)-1}).$$

Since $\#(\Pi_s) = l$, the term l cancels out and the coefficient of q^j equals $n_{j,s} - n_{j+1,s}$ for $j \geqslant 1$. On the other hand, the number of parts j in the dual partition also equals $n_{j,s} - n_{j+1,s}$. □

An easy verification shows that, for the root systems with two root lengths, the generalised exponents $e_1, \dots, e_l$ of the little adjoint representation are

$\mathbf{B}_n$ – n ($l = 1$); $\mathbf{C}_n$ – $2, 4, \dots, 2n-2$ ($l = n-1$); $\mathbf{F}_4$ – $4, 8$ ($l = 2$); $\mathbf{G}_2$ – 3 ($l = 1$).

In particular, if $e_1 \leqslant \dots \leqslant e_l$, then $e_i + e_{l+1-i} = h$ for all i.

Theorem 3.5. *For any* $\mu \in \Delta_s \cup \{0\}$*, the polynomial* $\mathfrak{M}^{\mu}_{\theta_s}(q)$ *depends only on* $\mathsf{ht}(\theta_s - \mu)$*, i.e., on* $\mathsf{ht}(\mu)$*. More precisely,*

- (i) $\mathfrak{M}^0_{\theta_s}(q) = q^{e_1} + \dots + q^{e_l}$;
- (ii) *If* $\mu \in \Delta_s^+$*, then* $\mathfrak{M}^{\mu}_{\theta_s}(q) = q^{\mathsf{ht}(\theta_s - \mu)}$;
- (iii) *if* $\alpha \in \Pi_s$*, then* $\mathfrak{M}^{-\alpha}_{\theta_s}(q) = (q-1)\mathfrak{M}^0_{\theta_s}(q) + q^{\mathsf{ht}(\theta_s)}$;
- (iv) *If* $\gamma \in \Delta_s^+$*, then* $\mathfrak{M}^{-\gamma}_{\theta}(q) = q^{\mathsf{ht}(\gamma)-1} \cdot \mathfrak{M}^{-\alpha}_{\theta_s}(q)$.

Proof. Part (i) is the subject of Lemma 3.4. The proof of other parts is similar to those of Theorem 3.1. □

Theorem 3.6. *We have* $\displaystyle\sum_{\mu \dashv \mathbb{V}_{\theta_s}} \mathfrak{M}^{\mu}_{\theta_s}(q) = \mathfrak{M}^0_{\theta_s}(q)(\mathfrak{M}^0_{\theta_s}(q) - l + 1) + \frac{\mathfrak{M}^0_{\theta}(q)}{q^{h-\mathsf{ht}(\theta_s)}} \cdot \frac{1-q^h}{1-q}$ *or, equivalently,*

$$\sum_{\mu \dashv \mathbb{V}_{\theta_s}} m^{\mu}_{\theta_s} \mathfrak{M}^{\mu}_{\theta_s}(q) = \mathfrak{M}^0_{\theta_s}(q)^2 + \frac{\mathfrak{M}^0_{\theta_s}(q)}{q^{h-\mathsf{ht}(\theta_s)}} \cdot \frac{1-q^h}{1-q}. \tag{8}$$

Proof. Our argument is similar to that of Theorem 3.3. Since $(e_l, e_{l-1}, \dots, e_1)$ and $(n_{1,s}, n_{2,s}, \dots)$ are dual partitions, we present Δ_s^+ as a union of l strings of roots, where the i-th string consists of roots of height $1, 2, \dots, e_i$. Then, using the fact that $e_j + e_{l-j+1} = h$ for all j, one computes that the sums of q-analogues of weight multiplicities over Δ_s^+ and $-\Delta_s^+$ are equal to $\dfrac{\mathfrak{M}^0_{\theta_s}(q) - lq^h}{q^{h-\mathsf{ht}(\theta_s)}(1-q)}$ and $((q-1)\mathfrak{M}^0_{\theta_s}(q) + q^{\mathsf{ht}(\theta_s)})\dfrac{l - \mathfrak{M}^0_{\theta_s}(q)}{(1-q)}$, respectively. □

Remark 3.7. It follows from this theorem that $\dfrac{\mathfrak{M}^0_{\theta_s}(q)}{q^{h-\mathsf{ht}(\theta_s)}}$ is a polynomial in q.

Remark 3.8. In the simply laced case, we have $l = r$, $\theta = \theta_s$, and $h - \mathsf{ht}(\theta_s) = 1$. Then Theorems 3.3 and 3.6 yield the same formulae.

Remark 3.9. Recall that the singular locus $\mathcal{N}^{sg}$ of $\mathcal{N}$ is irreducible (and the dense G-orbit in $\mathcal{N}^{sg}$ is said to be subregular). For any $\alpha \in \Pi_s$ and $\lambda \in Q \cap \mathfrak{X}_+$, Broer proves that the collapsing $G \times_{P_\alpha} \mathfrak{n}_\alpha \to \mathcal{N}^{sg}$, where $\mathfrak{n}_\alpha$ is the nilradical of Lie P_α, is birational and $\mathfrak{M}^0_\lambda(q) - q^{\mathsf{ht}(\alpha^+)}\mathfrak{M}^{\alpha^+}_\lambda(q)$ is the Poincaré polynomial counting the occurrences of $\mathbb{V}^*_\lambda$ in the graded ring $\Bbbk[\mathcal{N}^{sg}]$, i.e.,

$$\mathfrak{M}^0_\lambda(q) - q^{\mathsf{ht}(\alpha^+)}\mathfrak{M}^{\alpha^+}_\lambda(q) = \sum_i \dim_{\Bbbk} \mathsf{Hom}_G(\mathbb{V}^*_\lambda, \Bbbk[\mathcal{N}^{sg}]_i)\, q^i$$

[Br93, Corollary 4.7]. In particular, $m^0_\lambda - m^\alpha_\lambda$ is the multiplicity of $\mathbb{V}^*_\lambda$ in $\Bbbk[\mathcal{N}^{sg}]$. Using the Induction Lemma, we can prove that $q\,\mathfrak{M}^\alpha_\lambda(q) = q^{\mathsf{ht}(\alpha^+)}\mathfrak{M}^{\alpha^+}_\lambda(q)$. Therefore, this Poincaré polynomial is also equal to $\mathfrak{M}^0_\lambda(q) - q\,\mathfrak{M}^\alpha_\lambda(q) = q\,\mathfrak{M}^0_\lambda(q) - \mathfrak{M}^{-\alpha}_\lambda(q)$.

For the long simple root α, the collapsing $G \times_{P_\alpha} \mathfrak{n}_\alpha \to \mathcal{N}^{sg}$ is not birational and the ring $\Bbbk[\mathcal{N}^{sg}]$ should be replaced with $\Bbbk[G \times_{P_\alpha} \mathfrak{n}_\alpha]$.

4 A Weighted Sum of q-Analogues of All Weight Multiplicities

For a (possibly reducible) G-module $V = \sum_j a_j \mathbb{V}_{\lambda_j}$, we set $\mathfrak{M}^0_V(q) = \sum_j a_j \mathfrak{M}^0_{\lambda_j}(q)$. In [G87-2], R. Gupta (Brylinski) considered a $\mathbb{Z}[q]$-valued symmetric bilinear form $\langle\!\langle\ ,\ \rangle\!\rangle$ on the character ring of $\mathfrak{g}$:

$$\langle\!\langle \mathsf{ch}(V_1), \mathsf{ch}(V_2) \rangle\!\rangle = \mathfrak{M}^0_{V_1 \otimes V_2^*}(q).$$

She proved that this form has a nice expression via the q-analogues of *dominant* weights occurring in both V_1 and V_2. For any $\nu \in \mathfrak{X}_+$, consider the stabiliser $W_\nu \subset W$ and the restriction of the length function ℓ to W_ν. Set $t_\nu(q) = \sum_{w \in W_\nu} q^{\ell(w)}$, the Poincaré polynomial of W_ν. In particular, $t_0(q)$ is the Poincaré polynomial of W.

Theorem 4.1 ([G87-2, Corollary 2.4]). *For all* $\lambda, \gamma \in \mathfrak{X}_+$, *one has*

$$\langle\!\langle \mathsf{ch}(\mathbb{V}_\lambda), \mathsf{ch}(\mathbb{V}_\gamma)\rangle\!\rangle = \sum_{\nu\in\mathfrak{X}_+} \mathfrak{M}^\nu_\lambda(q)\mathfrak{M}^\nu_\gamma(q)\frac{t_0(q)}{t_\nu(q)}.$$

We provide here another formula for this bilinear form that involves the usual weight multiplicities for one representation and q-analogues of *all* weight multiplicities for the other representation. Below, we write $\mathfrak{M}^0_{\lambda^*\otimes\gamma}(q)$ in place of $\mathfrak{M}^0_{\mathbb{V}_{\lambda^*}\otimes\mathbb{V}_\gamma}(q)$.

Theorem 4.2. *For all* $\lambda, \gamma \in \mathfrak{X}_+$, *we have*

$$\sum_{\mu\dashv\mathbb{V}_\gamma} m^\mu_\gamma \mathfrak{M}^\mu_\lambda(q) = \sum_{\mu\dashv\mathbb{V}_\lambda} m^\mu_\lambda \mathfrak{M}^\mu_\gamma(q) = \mathfrak{M}^0_{\lambda^*\otimes\gamma}(q) = \sum_{\nu\in\mathfrak{X}_+} \mathfrak{M}^\nu_\gamma(q)\mathfrak{M}^\nu_\lambda(q)\cdot\frac{t_0(q)}{t_\nu(q)}.$$

Proof. The last equality here is the above-mentioned result of R. Brylinski; the first equality stems from the symmetry of the last expression with respect to λ and γ. Hence our task is to prove the second equality. Consider the vector bundle $G\times_B (\mathfrak{u}\oplus\mathbb{V}^*_\lambda)\to\mathcal{Z}$ and the corresponding sheaf $\mathcal{L}_\mathcal{Z}(\mathbb{V}^*_\lambda)$ of graded $\Bbbk[\mathcal{N}]$-modules. As in case of line bundles on $\mathcal{Z}$ (see Sect. 2), the graded character of $H^i(\mathcal{Z},\mathcal{L}_\mathcal{Z}(\mathbb{V}_\lambda)^\star)$ is well defined and we say that

$$\mathsf{gec}\,(\mathbb{V}_\lambda) = \sum_i (-1)^i \mathsf{ch}_q\big(H^i(\mathcal{Z},\mathcal{L}_\mathcal{Z}(\mathbb{V}_\lambda)^\star)\big)$$

is the *graded Euler characteristic* (of $\mathcal{L}_\mathcal{Z}(\mathbb{V}_\lambda)^\star$). Let us compute $\mathsf{gec}\,(\mathbb{V}_\lambda)$ in two different ways.

First, we can replace $\mathbb{V}_\lambda$ with the completely reducible B-module $\widetilde{\mathbb{V}}_\lambda = \oplus_\mu m^\mu_\lambda \Bbbk_\mu$, which does not change the graded Euler characteristic. Then

$$\mathsf{gec}\,(\mathbb{V}_\lambda) = \mathsf{gec}\,(\widetilde{\mathbb{V}}_\lambda) = \sum_{\mu\dashv\mathbb{V}_\lambda} m^\mu_\lambda\, \mathsf{gec}\,(\Bbbk_\mu) = \sum_{\mu\dashv\mathbb{V}_\lambda}\sum_{\nu\in\mathfrak{X}_+} m^\mu_\lambda \mathfrak{M}^\mu_\nu(q)\chi^*_\nu, \qquad (9)$$

where the last equality follows by Theorem 2.1.

On the other hand, $\mathbb{V}_\lambda$ is a G-module, therefore $G\times_B(\mathfrak{u}\oplus\mathbb{V}^*_\lambda)\simeq\mathcal{Z}\times\mathbb{V}^*_\lambda$ and

$$\mathsf{gec}\,(\mathbb{V}_\lambda) \simeq \mathsf{ch}(\mathbb{V}^*_\lambda)\cdot\mathsf{gec}\,(\Bbbk_0) = \chi^*_\lambda\cdot\sum_{\nu\in\mathfrak{X}_+}\mathfrak{M}^0_\nu(q)\chi^*_\nu. \qquad (10)$$

Now, equating the coefficients of χ^*_γ in (9) and (10), we will obtain the assertion. The required coefficient in (9) equals $\sum_{\mu\dashv\mathbb{V}_\lambda} m^\mu_\lambda\mathfrak{M}^\mu_\gamma(q)$. Expanding the product $\chi^*_\lambda\chi^*_\nu = \sum_{\kappa\in\mathfrak{X}_+} c^\kappa_{\lambda^*\nu^*}\chi_\kappa$, we see that the coefficient of χ^*_γ in (10) equals $\sum_{\nu\in\mathfrak{X}_+} c^{\gamma^*}_{\lambda^*\nu^*}\mathfrak{M}^0_\nu(q)$. Since $c^{\gamma^*}_{\lambda^*\nu^*} = c^\nu_{\lambda^*\gamma}$, this sum also equals $\mathfrak{M}^0_{\lambda^*\otimes\gamma}(q)$. □

Corollary 4.3. *For any* $\lambda \in \mathfrak{X}_+$, *we have* $\sum_{\mu \dashv \mathbb{V}_\lambda} m_\lambda^\mu \mathfrak{M}_\lambda^\mu(q) = \sum_{\nu \in \mathfrak{X}_+} \mathfrak{M}_\lambda^\nu(q)^2 \cdot \frac{t_0(q)}{t_\nu(q)}$.

(Note that for $\nu \in \mathfrak{X}_+$, $\mathfrak{M}_\lambda^\nu(q)$ is nonzero if and only if $\nu \dashv \mathbb{V}_\lambda$.) This equality shows that the weighted sum $\sum_{\mu \dashv \mathbb{V}_\lambda} m_\lambda^\mu \mathfrak{M}_\lambda^\mu(q)$ is a more natural object than just $\sum_{\mu \dashv \mathbb{V}_\lambda} \mathfrak{M}_\lambda^\mu(q)$. Actually, we do not know any closed expression for the latter. Moreover, the weighted sum of q-analogues of all weight multiplicities is a polynomial with nonnegative coefficients, whereas this is not always the case for the plain sum (use Theorem 3.3 and look at the adjoint representation of $\mathfrak{sl}_{r+1}$ with $r \geqslant 4$).

Remark 4.4. It was tempting to conjecture that Corollary 4.3 could be refined so that one takes the sum over a sole Weyl group orbit $W\mu$ in the LHS and pick the summand corresponding to μ^+ in the RHS. But this doesn't work! For instance, if $\mu = 0 \dashv \mathbb{V}_\lambda$, then the corresponding summands are $m_\lambda^0 \mathfrak{M}_\lambda^0(q)$ (left) and $\mathfrak{M}_\lambda^0(q)^2$ (right).

Corollary 4.5. *1) If* $\lambda \in \mathfrak{X}_+$ *is minuscule, then* $\sum_{\mu \dashv \mathbb{V}_\lambda} \mathfrak{M}_\lambda^\mu(q) = \sum_{\mu \dashv \mathbb{V}_\lambda} q^{\mathsf{ht}(\lambda-\mu)} = \frac{t_0(q)}{t_\lambda(q)}$;

2) More generally, if $\mathbb{V}_\lambda$ *is weight multiplicity free (i.e.,* $m_\lambda^\mu = 1$ *for all* $\mu \dashv \mathbb{V}_\lambda$*), then*

$$\sum_{\mu \dashv \mathbb{V}_\lambda} \mathfrak{M}_\lambda^\mu(q) = \sum_{\mu \dashv \mathbb{V}_\lambda} q^{\mathsf{ht}(\lambda-\mu)}.$$

Proof. 1) In this case all weight multiplicities are equal to one and λ is the only dominant weight of $\mathbb{V}_\lambda$. Moreover, all the weights μ satisfy the condition (3) of Theorem 2.2 and therefore $\mathfrak{M}_\lambda^\mu(q) = q^{\mathsf{ht}(\lambda-\mu)}$.

2) Since $m_\lambda^\mu = 1$ for all μ, using Theorem 2.4, one easily proves by induction on $\mathsf{ht}(\lambda - \mu)$ that $\mathfrak{M}_\lambda^\mu(q) = q^{\mathsf{ht}(\lambda-\mu)}$.

□

This corollary shows that, for the weight multiplicity free case, $\sum_{\mu \dashv \mathbb{V}_\lambda} \mathfrak{M}_\lambda^\mu(q)$ equals the *Dynkin polynomial* of $\mathbb{V}_\lambda$, see [P04, Sect. 3].

Corollary 4.6. *(i) If* $\mathfrak{g}$ *is simply laced, then* $\frac{t_0(q)}{t_\theta(q)} = \frac{\mathfrak{M}_\theta^0(q)}{q} \cdot \frac{1-q^h}{1-q}$.

(ii) More generally, for any simple Lie algebra $\mathfrak{g}$, *we have* $\frac{t_0(q)}{t_{\theta_s}(q)} = \frac{\mathfrak{M}_{\theta_s}^0(q)}{q^{h-\mathsf{ht}(\theta_s)}} \cdot \frac{1-q^h}{1-q}$ *and* $\frac{t_0(q)}{t_\theta(q)} = \frac{\mathfrak{M}_{\theta^\vee}^0(q)}{q^{h-\mathsf{ht}(\theta^\vee)}} \cdot \frac{1-q^h}{1-q}$, *where* $\theta^\vee$ *is regarded as the short dominant root in* $\Delta^\vee$.

Proof. In the simply laced case, $\theta = \theta_s$ and $\mathsf{ht}(\theta_s) = \mathsf{ht}(\theta^\vee) = h - 1$. Therefore, it suffices to prove part (ii). For the first equality in (ii), we combine Eq. (8) and Corollary 4.3 with $\lambda = \theta_s$, and also use the fact that the only dominant weights of $\mathbb{V}_{\theta_s}$ are θ_s and 0. The second equality stems from the similar argument for the dual Lie algebra $\mathfrak{g}^\vee$ and the fact that $t_\theta(q) = t_{\theta^\vee}(q)$. □

Remark 4.7. The last corollary can be verified by a direct calculation. Recall that if $d_1(\nu), \dots, d_r(\nu)$ are the degrees of basic invariants of the reflection group $W_\nu \subset GL(\mathfrak{t})$, then $t_\nu(q) = \prod_{i=1}^{r} \dfrac{1-q^{d_i(\nu)}}{1-q}$. In particular, $d_i(0) = m_i + 1$. It is a kind of miracle that the complicated fraction $\dfrac{t_0(q)}{t_{\theta_s}(q)} = \prod_{i=1}^{r} \dfrac{1-q^{d_i(0)}}{1-q^{d_i(\theta_s)}}$ simplifies to rather a simple expression!

5 Generalised Exponents and the Height of Weights

In Sect. 3, we used the fact that the generalised exponents of $\mathfrak{g} = \mathbb{V}_\theta$ and $\mathbb{V}_{\theta_s}$ are determined via the height of "positive weights (roots)". Here we provide a geometric condition for this phenomenon and point out some other irreducible representation having the similar property. This relies on results of R. Brylinski on the principal filtration of a weight space and "jump" polynomials [B89].

Let e be a principal nilpotent element of $\mathfrak{g}$ and $\{e, \tilde{h}, f\}$ a corresponding principal $\mathfrak{sl}_2$-triple in $\mathfrak{g}$. Without loss of generality, we assume that e is the sum of root vectors corresponding to Π and hence $\alpha(\tilde{h}) = 2$ for all $\alpha \in \Pi$ [Ko59, Ko63]. This means that upon the identification of $\mathfrak{t}$ and $\mathfrak{t}^*$, $\frac{1}{2}\tilde{h}$ is nothing but $\rho^\vee := \frac{1}{2}\sum_{\gamma\in\Delta^+} \gamma^\vee$ and $\gamma(\frac{1}{2}\tilde{h}) = (\gamma, \rho^\vee) = \mathsf{ht}(\gamma)$ for all $\gamma \in Q_+$.

Let $\mathfrak{s} = \langle e, \tilde{h}, f\rangle$ be the corresponding simple subalgebra of $\mathfrak{g}$. We write $\mathcal{R}_n$ for the simple $\mathfrak{s}$-module of dimension $n+1$, so that the $\tilde{h}$-eigenvalues in $\mathcal{R}_n$ are $n, n-2, \dots, -n$.

In what follows, $\mathbb{V} = \mathbb{V}_\lambda$, $\mathcal{P}(\mathbb{V})$ is the set of weights of $\mathbb{V}$, and

$$\mathcal{P}(\mathbb{V})_+ = \{\nu \in \mathcal{P}(\mathbb{V}) \mid \nu(\tilde{h}) > 0\}.$$

We also write $\widetilde{\mathcal{P}(\mathbb{V})_+}$ for the *multiset* of weights in $\mathcal{P}(\mathbb{V})_+$, where each ν appears with multiplicity m_λ^ν. It is assumed below that $\lambda \in \mathfrak{X}_+ \cap Q$, so that $\mathcal{P}(\mathbb{V}) \subset Q$ and $m_\lambda^0 \neq 0$.

Theorem 5.1. *Suppose that* $\dim \mathbb{V}^{\mathfrak{t}} = \dim \mathbb{V}^e (=: n)$. *Then*

(i) $\mathcal{P}(\mathbb{V}) = \mathcal{P}(\mathbb{V})_+ \cup \{0\} \cup (-\mathcal{P}(\mathbb{V})_+)$*; moreover, each nonzero weight is a multiple of a root.*

(ii) $$\prod_{\gamma\in\widetilde{\mathcal{P}(\mathbb{V})_+}} \frac{1-q^{\mathsf{ht}(\gamma)+1}}{1-q^{\mathsf{ht}(\gamma)}} = \prod_{i=1}^{n} \frac{1-q^{m_i(\lambda)+1}}{1-q}.$$

Proof. (i) Considering $\mathbb{V}$ as $\mathfrak{s}$-module, we obtain a decomposition $\mathbb{V}|_{\mathfrak{s}} = \bigoplus_{j=1}^{n} \mathcal{R}_{l_j}$. Since $\dim \mathbb{V}^e \geqslant \dim \mathbb{V}^{\tilde{h}} \geqslant \dim \mathbb{V}^{\mathfrak{t}}$, the hypothesis implies that each $\mathcal{R}_{l_j}$ has a zero-weight space (hence each $l_j = 2k_j$ is even) and $\mathbb{V}^{\mathfrak{t}} = \mathbb{V}^{\tilde{h}}$. Consequently, if $\nu \in \mathcal{P}(\mathbb{V})$ and $\nu \neq 0$, then $\nu(\tilde{h}) \neq 0$, which proves the partition formula. Letting $\mathbb{V}_{\pm} = \oplus_{\gamma\in\mathcal{P}(\mathbb{V})_{\pm}} \mathbb{V}^{\gamma}$, we see that $\mathbb{V}^{\mathfrak{t}}$ generates $\mathbb{V}^{\mathfrak{t}} \oplus \mathbb{V}_+$ as e-module, whence $\mathcal{P}(\mathbb{V})_+ \subset Q_+$. Now, it is easily seen that if $\gamma \in Q_+$ is not proportional to a root, then there exists $w \in W$ such that $w(\gamma) \notin Q_+ \cup (-Q_+)$.

(ii) By the above decomposition of $\mathbb{V}|_{\mathfrak{s}}$, the multiset of positive $\frac{1}{2}\tilde{h}$-eigenvalues in $\mathbb{V}$, i.e., the multiset $\{\mathsf{ht}(\gamma) \mid \gamma \in \widetilde{\mathcal{P}(\mathbb{V})_+}\}$ consists of $\{1, 2, \dots, k_1,\ 1, 2, \dots, k_2,\ \dots,\ 1, 2, \dots, k_n\}$. Therefore, most of the factors cancel out in the LHS and we are left with the product $\prod_{j=1}^{n} \frac{1-q^{k_i+1}}{1-q}$.

On the other hand, the theory of R. Brylinski [B89] shows that the generalised exponents of $\mathbb{V}$ are determined by the e-filtration on $\mathbb{V}^{\mathfrak{t}}$ and are equal to the $\frac{1}{2}\tilde{h}$-eigenvalues in $V^{\mathfrak{z}(e)}$, where $\mathfrak{z}(e)$ is the centraliser of e in $\mathfrak{g}$. Since $\dim \mathbb{V}^{\mathfrak{z}(e)} = \dim \mathbb{V}^{\mathfrak{t}}$ for the simple G-modules having zero weight [B89, Corollary 2.7], it follows that $\mathbb{V}^{\mathfrak{z}(e)} = \mathbb{V}^e$ in our situation, and the eigenvalues in question are $k_1, k_2, \dots, k_n$. This yields the desired equality in part (ii). □

Remark 5.2. A formal consequence of relation (ii) is that the partition $\big(m_1(\lambda), \dots, m_n(\lambda)\big)$ is dual to the partition formed by the numbers $\#\{\gamma \in \widetilde{\mathcal{P}(\mathbb{V})_+} \mid \mathsf{ht}(\gamma) = i\}$. For $\mathbb{V} = \mathfrak{g}$ and $\mathcal{P}(\mathbb{V})_+ = \Delta^+$, formula (ii) is sometimes called the Kostant-Macdonald identity, see [AC89]; we also refer to [AC12] for a recent generalisation related to Schubert varieties.

Remark 5.3. If $\mathbb{V}$ is a simple G-module with non-trivial zero-weight space, then

$$\dim \mathbb{V}^e \geqslant \dim \mathbb{V}^{\mathfrak{z}(e)} = \dim \mathbb{V}^{\mathfrak{t}} \leqslant \dim \mathbb{V}^{\tilde{h}}$$

and $\dim \mathbb{V}^e \geqslant \dim \mathbb{V}^{\tilde{h}}$. Therefore the hypothesis of Theorem 5.1 implies that all these spaces have one and the same dimension. It is also known that, for *any* simple G-module $\mathbb{V}$, $\dim \mathbb{V}^{\mathfrak{z}(e)}$ equals the dimension of a largest weight space, which is achieved for either the unique dominant minuscule weight or zero, see [gr92, Remark 1.6].

Making use of the above coincidences and Theorem 5.1(i), one easily proves that the hypothesis of the theorem holds exactly for the following pairs $(\mathfrak{g}, \lambda)$ with simple $\mathfrak{g}$:

- $(\mathfrak{g}, \theta)$ and $(\mathfrak{g}, \theta_s)$, i.e., the adjoint and little adjoint representations of $\mathfrak{g}$;
- $(\mathbf{B}_r, 2\varphi_1)$, $(\mathbf{G}_2, 2\varphi_1)$, $(\mathbf{A}_1, 2m\varphi_1)$, $m \in \mathbb{N}$.

The generalised exponents for the first two cases in the second line are $2, 4, \dots, 2r$ and $2, 4, 6$, respectively.

Acknowledgements Part of this work was done while I was visiting the Friedrich-Schiller-Universität (Jena) in Spring 2013.

References

[AC89] E. Akyildiz, J.B. Carrell, A generalization of the Kostant-Macdonald identity. Proc. Natl. Acad. Sci. (USA) **86**, 3934–3937 (1989)

[AC12] E. Akyildiz, J.B. Carrell, Betti numbers of smooth Schubert varieties and the remarkable formula of Kostant, Macdonald, Shapiro, and Steinberg. Mich. Math. J. **61**, 543–553 (2012)

[Ba01] Yu. Bazlov, Graded multiplicities in the exterior algebra. Adv. Math. **158**, 129–153 (2001)

[Br93] A. Broer, Line bundles on the cotangent bundle of the flag variety. Invent. Math. **113**, 1–20 (1993)

[br94] A. Broer, Normality of some nilpotent varieties and cohomology of line bundles on the cotangent bundle of the flag variety, in *Lie Theory and Geometry*, ed. by J.-L. Brylinski et al. Progress in Mathematics, vol. 123 (Boston, Birkhäuser, 1994), pp. 1–19

[Br97] A. Broer, A vanishing theorem for Dolbeault cohomology of homogeneous vector bundles. J. Reine Angew. Math. **493**, 153–169 (1997)

[B89] R.K. Brylinski, Limits of weight spaces, Lusztig's q-analogs, and fiberings of adjoint orbits. J. Am. Math. Soc. **2**, 517–533 (1989)

[gr92] W.A. Graham, Functions on the universal cover of the principal nilpotent orbit. Invent. Math. **108**, 15–27 (1992)

[G87-1] R.K. Gupta, Characters and the q-analog of weight multiplicity. J. Lond. Math. Soc. **36**, 68–76 (1987)

[G87-2] R.K. Gupta, Generalized exponents via Hall-Littlewood symmetric functions. Bull. Am. Math. Soc. **16**, 287–291 (1987)

[He76] W. Hesselink, Cohomology and the resolution of the nilpotent variety. Math. Ann. **223**(3), 249–252 (1976)

[He80] W. Hesselink, Characters of the nullcone. Math. Ann. **252**, 179–182 (1980)

[Ion04] B. Ion, The Cherednik kernel and generalized exponents. IMRN **2004**(36), 1869–1895 (2004)

[Ka82] S. Kato, Spherical functions and a q-analogue of Kostant's weight multiplicity formula. Invent. Math. **66**, 461–468 (1982)

[Ko59] B. Kostant, The principal three-dimensional subgroup and the Betti numbers of a complex simple Lie group. Am. J. Math. **81**, 973–1032 (1959)

[Ko63] B. Kostant, Lie group representations on polynomial rings. Am. J. Math. **85**, 327–404 (1963)

[Lu83] G. Lusztig, Singularities, character formulas, and a q-analog of weight multiplicities. *Analyse et Topologie Sur Les Espaces Singuliers* (II–III). Astérisque, vol. 101–102 (Société Mathématique de France, Paris, 1983), pp. 208–227

[P04] D. Panyushev, Weight multiplicity free representations, $\mathfrak{g}$-endomorphism algebras, and Dynkin polynomials. J. Lond. Math. Soc. **69**, Part 2, 273–290 (2004)

[P10] D. Panyushev, Generalised Kostka-Foulkes polynomials and cohomology of line bundles on homogeneous vector bundles. Sel. Math. (New Series) **16**, 315–342 (2010)
[S90] H. Samelson, *Notes on Lie Algebras*. Universitext, 2nd edn. (Springer, New York, 1990), xii+162 pp.
[V06] S. Viswanath, A note on exponents vs root heights for complex simple Lie algebras. Electr. J. Comb. **13**, # N22, 5 pp. (2006)

The Triangulation of Manifolds: Topology, Gauge Theory, and History

Frank Quinn

Abstract A mostly expository account of old questions about the relationship between polyhedra and topological manifolds. Topics are old topological results, new gauge theory results (with speculations about next directions), and history of the questions.

MR classifications 57Q15, 01A60, 57R58.

1 Introduction

This is a survey of the current state of triangulation questions posed by Kneser in 1924:

(1) Is a polyhedron with the local homology properties of Euclidean space, locally homeomorphic to Euclidean space?
(2) Is a space locally homeomorphic to Euclidean space, triangulable (homeomorphic to some polyhedron)?
(3) If there are two such triangulations, must they be PL equivalent?

Topological work on the topic is described in Sects. 2–3. This work was mature and essentially complete by 1980, but leaves open questions about H-cobordism classes of homology 3-spheres. Gauge theory has had some success with these, with the most substantial progress for the triangulation questions made in a recent paper of Manolescu [mano13]. Manolescu's paper is discussed in Sect. 4. This area is not yet mature, and one objective is to suggest other perspectives.

Section 5 recounts some of the history of Kneser's questions. Kneser posed them as an attempt to provide foundations for Poincaré's insights 20 years before, but

F. Quinn (✉)
Department of Mathematics, Virginia Polytechnic Institute and State University, 476 McBryde Hall, Blacksburg, VA 24061-0123, USA
e-mail: quinn@math.vt.edu

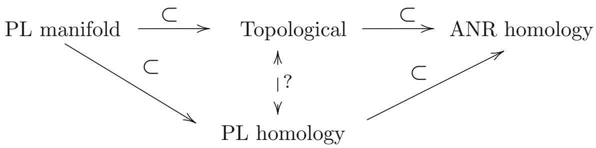
W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_11

they turned out to be a dead end. They were fruitful challenges to technology as it developed, and much of the progress of the subject can be traced out in applications to these questions, but they do not themselves have significant applications. Here we have a different concern: why did Kneser point his contemporaries into a dead end? Or was he trying to get them to face the fact that it was a dead end? The question gives a fascinating window into the transition from pre-modern to modern mathematics in the early twentieth century.

The remainder of the introduction gives modern context for the questions and describes the organization of the technical parts of the paper.

1.1 Modern Context

The relevant main-line topics are PL manifolds, topological manifolds, and ANR homology manifolds.[1] Polyhedra that are homology manifolds, referred to here as "PL homology manifolds," are mixed-category objects, and Kneser's questions amount to asking how these are related to the main-line categories.



The standard categories differ radically in flavor and technique, but turn out to be almost equivalent. For the purposes here, topological and ANR homology manifolds are equivalent.[2] PL and topological manifolds differ by the Kirby–Siebenmann invariant $\mathsf{ksm}(M) \in H^4(M; Z/2)$. This is in a single cohomology group, with the smallest possible coefficients, so is about as small as an obstruction can be without actually being zero. This means the *image* of PL homology manifolds in the main-line picture is highly constrained, and on the image level the answers to the questions can't be much different from "yes." Unfortunately it turns out that there are a great many PL homology manifolds in each image equivalence class.

[1] ANR = "Absolute Neighborhood Retract." For finite-dimensional spaces this is equivalent to "locally contractible," and is used to rule out local point-set pathology.

[2] There are "exotic" homology manifolds not equivalent to topological manifolds, [BFMW] but they are extremely difficult to construct and are not produced by any known natural process.

1.2 Topology

PL homology manifolds have two types of singularities: dimension 0 (problematic links of vertices) and codimension 4 (problematic links of $(n-4)$-simplices). Four-manifolds are special, in part because these two types coincide. Vertex singularities can be canonically resolved, so are topologically inessential. This is described in Sect. 2 and provides an answer to Kneser's first question.

The real difficulties come from the codimension 4 singularities, and these involve homology spheres. We denote the group of homology H-cobordism classes of homology 3-spheres by Θ. The full official name is Θ_3^H, but the decorations are omitted here because they don't change. Several descriptions of this group are given in Sect. 3.1.

A PL homology manifold K has an easily defined Cohen–Saito–Sullivan cohomology class $\mathsf{css}(K) \in H^4(K;\Theta)$ [sullSing, cohen]; see Sect. 3.6. The Rokhlin homomorphism $\mathsf{rok}\colon \Theta \to Z/2$ induces a change-of-coefficients exact sequence

$$\longrightarrow H^4(K;\ker(\mathsf{rok})) \longrightarrow H^4(K;\Theta) \xrightarrow{\mathsf{rok}} H^4(K;Z/2) \xrightarrow{\beta} H^5(K;\ker(\mathsf{rok}))$$

with Bokstein connecting homomorphism β. The image of the Cohen–Saito–Sullivan class is the Kirby–Siebenmann class. The baby version of the main theorem is

Theorem 1 ([gs]). *If M is a topological manifold of dimension $\neq 4$ (and boundary of dimension $\neq 4$ if it is nonempty), then concordance classes of homeomorphisms to polyhedra correspond to lifts of the Kirby–Siebenmann class to $H^4(M;\Theta)$.*

This is the baby version because serious applications (if there were any) would require the relative version, Theorem 3.1. A corollary is that a triangulation exists if and only if the Bokstein of the Kirby–Siebenmann invariant is trivial. Note that if the Kirby–Siebenmann class lifts to a class with integer coefficients, then it lifts to any coefficient group, and it follows that the manifold is triangulable. Similarly, if the Kirby–Siebenmann class lifts to coefficients Z/k but no further, then triangulability of the manifold depends on whether or not there is an element in Θ of order k and nontrivial Rokhlin invariant. Finally, triangulations are classified up to concordance by $H^4(K;\ker(\mathsf{rok}))$.

These results reduce the geometric questions to questions about the group Θ and the Rokhlin homomorphism. This part of the picture was essentially complete by 1980, but Θ is opaque to traditional topological methods. It has grudgingly yielded some of its secrets to sophisticated gauge theory; an overview is given in Sect. 4. It is infinitely generated and lots of these generators have infinite order. This means if there is a triangulation of M and $H^4(M;Z) \neq 0$, then there are a great many different ones. Manolescu's recent advance is that the Rokhlin homomorphism does not split. This implies that there are manifolds (e.g., the ones identified by Galewski–Stern

[gs5]) that cannot be triangulated. Manolescu's paper is described in Sect. 4. This theory is in a relatively early stage of development so the section gives speculations about future directions.

It seems reasonable to speculate that homology spheres with nontrivial Rokhlin invariant must have infinite order. Indeed, it seems reasonable to expect Θ to be torsion-free. Either would imply that M has a triangulation if and only if the integral Bokstein $\beta: H^4(M; Z/2) \to H^5(M; Z)$ is trivial on the Kirby–Siebenmann class. Proof of existence of triangulations in such cases should be easier than cases that might involve torsion.

2 Homology Manifolds are Essentially Manifolds

We begin with Kneser's first question because the answer is easy (now) and sets the stage for the others.

2.1 Singular Vertices

Suppose L is a PL homology n-manifold with homology isomorphic to the homology of the n-sphere. Then the cone on L is a PL homology $(n+1)$-manifold (with boundary). However if the dimension is greater than 1 (to exclude circles) and L is *not simply connected*, then the cone point is not a manifold point. The reason is that the relative homotopy group

$$\pi_2(\text{cone}, \text{cone} - *) \simeq \pi_1(L) \neq \{1\}$$

is nontrivial, and this is impossible for a point in a manifold.

The lowest dimension in which non-simply connected homology spheres occur is 3, and the oldest and most famous 3-dimensional example was described by Poincaré, see Kirby–Scharlemann [kscharlemann79]. There are examples in all higher dimensions but the 3-dimensional ones are the most problematic. These cone points turn out to be the only topological singularities:

Theorem 2.1. *A PL homology manifold is a topological manifold except at vertices with non-simply connected links of dimension greater than 2.*

This statement is for manifolds without boundary, but extends easily. Boundary point is singular if either the link in the boundary, or the link in the whole manifold, is non-simply connected.

We give a quick proof using mature tools from the study of ANR homology manifolds. Most homology manifolds are not manifolds, and some of them are quite ghastly.[3] Nonetheless they are close to being manifolds. There is a single obstruction in $H^0(X; Z)$ whose vanishing corresponds to the existence of a map $M \to X$ with essentially contractible point inverses, and M a topological manifold [Q1]. These are called *resolutions* by analogy with resolution of singularities in algebraic geometry. When a resolution exists it is unique, essentially up to homeomorphism. Roughly speaking this gives an equivalence of categories, and the global theories are the same.

The obstruction is so robust that a heroic effort was required to show that exotic examples exist [BFMW]. Existence of a manifold point implies the obstruction vanishes, so PL homology manifolds have resolutions.

Next, Edwards' CE approximation theorem asserts that if X is an ANR homology manifold of dimension at least 5, and $r: M \to X$ is a resolution, then r can be approximated by a homeomorphism if and only if X has the "disjoint 2-disk property," see [daverman]. It is easy to see that PL homology manifolds of dimension at least 5 have the disjoint disk property everywhere except at π_1-bad vertices. This completes the proof except in dimension 4, where the only question is with cones on homotopy spheres. Perelman has shown that these are actually standard, so the cone is a PL 4-ball and the cone point is a PL manifold point. The weaker assertion that they are topologically standard also follows from the next section.

This proof seems effortless because we are using big hammers on small nails. The job could be done with much smaller hammers, but this is more complicated and might give the impression that we don't have big hammers. Also, as mentioned in the introduction, there is a rich history of partial results not recounted here.

2.2 Resolutions with Collared Singularities

The proof given above uses the fact that singularities in ANR homology manifolds can be "resolved." The next theorem gives a precise refinement for the PL case, based on the following lemma:

Lemma 2.2. *Suppose L is a PL homology manifold with the homology of a sphere. Then L bounds a contractible manifold in the sense that there is a contractible ANR homology manifold W with $\partial W = L$, L has a collar neighborhood in W, and $W - L$ is a topological manifold. Further, any two such W are homeomorphic rel a neighborhood of the boundary.*

The only novelty is that we have not assumed L is a manifold. The proof of the 4-dimensional case given in [fq, Corollary 9.3C] extends easily. We sketch the proof.

[3] There is a technical definition of "ghastly" in [davermanwalsh] that lives up to the name.

The standard triangulation of $L \times [0, 1]$ has no vertices in the interior so, by the Theorem above, the interior is a manifold. Do the plus construction [fq, §11.1] to kill the fundamental group. The result is M with manifold interior, collared boundary $L \times [0, 1]$, and proper homotopy equivalent to a sphere. Replace each $L \times [n, n+1] \subset L \times [0, \infty)$ by a copy of M and denote the result by W. If W is a manifold except at the singularities of L, then the standard manifold proof shows that the 1-point compactification is contractible, and a manifold except for these same singularities. It also shows that this manifold is unique up to homeomorphism rel boundary. The modification required in the older proof is verification that the interior of W is a manifold.

W is a manifold except possibly at vertices in $L \times \{n\}$ where the copies are glued together. If $n > 0$, then $L \times \{n\}$ has a collar on each side, so has a neighborhood homeomorphic to $L \times \mathbf{R}$, which is a manifold. Thus the only non-manifold points are in $L \times \{0\}$. This completes the proof.

We use the lemma to define models for "collared singular points." Suppose W is as in the lemma, with boundary collar $L \times [0, 1) \to W$. Identify the complement of a smaller open collar to a point to get $W \to W/(W - L \times [0, 1/2))$. The quotient is the cone $L \times [0, 1/2]/(L \times \{1/2\})$, the map is a homeomorphism except at the cone point, and the preimage of this point is a smaller copy of W and therefore contractible. In particular this is a resolution.

Now define a "resolution with collared singular points" to be $M \to K$, that is, a homeomorphism except at a discrete set of points in K, and near each of these points is equivalent to a standard model. The lemma easily implies:

Theorem 2.3. *A PL homology manifold K has a topological resolution with collared singular points, and singular images the π_1-bad vertices of K. This resolution is well-defined up to homeomorphism commuting with the maps to K.*

The mapping cylinder of a resolution is a homology manifold, and can be thought of as a "concordance" between domain and range. In these terms the theorem asserts that a PL homology manifold is concordant in a strong sense to a manifold.

The unusually strong uniqueness (commuting exactly with maps to K, not just arbitrarily close) results from the fact that two such resolutions have the same singular images, and the uniqueness in Lemma 2.2. This statement is true for manifolds with boundary if the definition of "collared singularity" is extended in the straightforward way.

3 Triangulation

The main theorem is stated after the obstruction group is defined. The proof has two parts: first, enough structure of homology manifolds is developed to see the Cohen–Saito–Sullivan invariant. Both cohomology and homology versions are described, in part to clarify the role of orientations. The second part is the converse, due to Galewski and Stern.

3.1 The Group

Θ is usually defined as the set of oriented homology 3-spheres modulo homology H-cobordism. Connected sum defines an abelian monoid structure, and this is a group because reversing orientation gives additive inverses. As mentioned in the introduction, the full name of this group is Θ_3^H, but analogous groups Θ_k^H for $k \geq 3$ are, fortunately, trivial. Roughly speaking, nontriviality would come from fundamental groups, and in higher dimensions we can kill these (e.g., with plus constructions).

Geometric constructions give disjoint unions of homology spheres, not single spheres. These can be joined by connected sum to give an element in the usual definition of the group, but there are a number of advantages to using a definition that accepts disjoint unions directly. In this view Θ is a quotient of the free abelian group generated by homology 3-spheres. Elements in the kernel are boundaries of oriented PL four-manifolds that are homologically like D^4 minus the interiors of finitely many disjoint 4-balls. These boundaries are disjoint unions of homology 3-spheres, and we identify disjoint unions with formal sums in the abelian group. Elements of the standard version are generators in the expanded version. It is an easy exercise to see that this inclusion gives an isomorphism of groups.

In either definition it is important that the equivalence relation be defined by PL manifolds, not just homology manifolds. The goal is to organize $\mathsf{cone}(L)$-type singularities, and allowing singularities in the equivalences would defeat this. There may eventually be applications in which "concordances" can have limited singularities and the corresponding obstruction group should have these singularities factored out. For instance, Gromov limits of Riemannian manifolds with special metrics might allow variation by cones on homology spheres with special metrics.

The Rokhlin invariant is a homomorphism $\mathsf{rok}\colon \Theta \to Z/2$ defined using signatures of spin four-manifolds bounding homology 3-spheres, cf., [kirby]. This connects with the Kirby–Siebenmann invariant, as described next.

Theorem 3.1 (Main Theorem).

(1) (CSS invariant) *A PL homology manifold K has a "Cohen–Saito–Sullivan" invariant* $\mathsf{css}(K) \in H^4(K; \Theta)$*;*

(2) (Relation to Kirby–Siebenmann) *If $r\colon M \to K$ is a topological resolution of a PL homology manifold, then* $\mathsf{ksm}(M) = \mathsf{rok}(r^*(\mathsf{css}(K)))$*; and*

(3) (Realization: Galewski–Stern [gs]) *Suppose M is a topological manifold, not dimension 4, and a homeomorphism $\partial M \to L$ to a polyhedron is given. If there is a lift ℓ of* $\mathsf{ksm}(M)$ *to Θ that extends* $\mathsf{css}(L)$*, then there is a polyhedral pair (K, L) and a homeomorphism $M \to K$ that extends the homeomorphism on ∂M, and* $\mathsf{css}(K) = \ell$.

Here, a "lift" is an element ℓ:

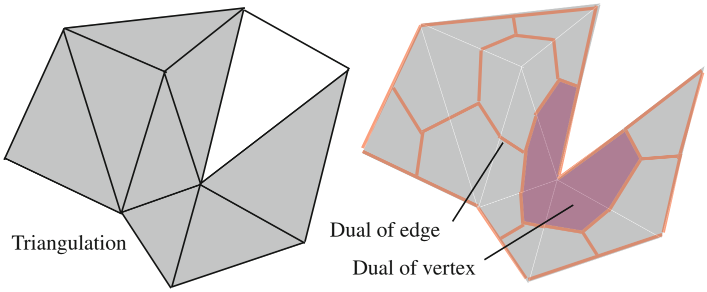


Fig. 1 Dual cones in a simplicial complex

$$\begin{array}{ccc} \ell & \longrightarrow & \mathsf{css}(\partial K) \\ \downarrow & & \downarrow \\ \mathsf{ksm}(M) & \longrightarrow & \mathsf{ksm}(\partial M) \end{array} \quad \text{in} \quad \begin{array}{ccc} H^3(M;\Theta) & \xrightarrow{\partial *} & H^3(\partial K;\Theta) \\ \downarrow \mathsf{rok} & & \downarrow \mathsf{rok} \\ H^3(M;Z/2) & \xrightarrow{\partial *} & H^3(\partial K;Z/2) \end{array}$$

There is a slightly sharper version in which "lift" is interpreted as a cochain representing such a cohomology class. Another extension is that if the map $\partial M \to L$ is a resolution instead of a homeomorphism, then the conclusion is that it extends to $M \to K$, that is, a homeomorphism on $M - \partial M$. The significance is that vertex singularities in L (where $\partial M \to L$ cannot be a homeomorphism) do not affect the codimension-4 obstructions. Finally, the fact that four-manifolds are smoothable in the complement of points [ends3] can be used to alter definitions to give a formulation that includes dimension 4. We await guidance from applications to see which of these refinements is worth writing out.

The proof of parts (1) and (2) are given in the remainder of this section. The Galewski–Stern proof of (3) follows the pattern developed to classify smooth and PL structures [ks], so is more elaborate than really needed. I did not find a proof short enough to include here, however.

3.2 Structure of Polyhedra

We review the structure of polyhedra needed for homology manifolds. Suppose σ is a simplex in a simplicial complex. The *dual cone* of σ is a subcomplex of the barycentric subdivision of the complex. Specifically, it is the collection of simplices that intersect σ in exactly the barycenter. The *link* is the subcomplex of this consisting of faces opposite to the barycenter point (Fig.1).

It is easy to see that the dual cone is the cone on the link, with cone point the barycenter of σ. This extends to an embedding of the join of the link and σ, using linearity in the simplex. Here we only need the weaker conclusion that the interior of σ has a neighborhood isomorphic to the product $\mathsf{int}(\sigma) \times \mathsf{cone}(\mathsf{link}(\sigma))$.

3.3 Links in PL Homology Manifolds

Recall that X is a homology n-manifold (without boundary) if for each $x \in X$, $H_*(X, X-x; \mathbf{Z}) \simeq H_*(\mathbf{R}^n, \mathbf{R}^n - o; \mathbf{Z})$. A pair is a homology manifold with boundary if $X - \partial X$ is a homology n-manifold, ∂X is a homology $(n-1)$-manifold, and points in the boundary have the same local homology as points in the boundary of an n-ball (i.e., trivial).

PL homology manifolds have much more structure.

Lemma 3.2. *A polyhedron K is a homology n-manifold (without boundary) if and only if link of every simplex is a homology manifold, and has the homology of an $(n-k-1)$-sphere, where k is the dimension of the simplex.*

The statement about homology of links is an easy suspension argument. The assertion that links are homology manifolds follows from this and the fact that links in a link also appear as links in the whole space (easy after unwinding definitions). This statement is easily extended to a version for manifolds with boundary.

3.6 The Cohomology Picture

We begin with the cohomological version of the codimension-4 invariant.

In a homology manifold the cones have the relative homology of disks, so they give a model for the chain complex. Specifically, define the *conical chain* group $C_n^{\mathsf{cone}}(K)$ to be the free abelian group generated by n-dimensional cones together with a choice of orientation. Boundary homomorphisms in this complex come from homology exact sequences in a standard way.

Define a homomorphism Z[oriented 4-d cones] $\to \Theta$ by

$$(\mathsf{cone}(L), \alpha) \mapsto [L, \partial\alpha]$$

where α denotes the orientation of the (4-dimensional) cone, and $\partial\alpha$ the corresponding orientation of the (3-dimensional) homology sphere. It is not hard to see that this defines a cohomology class [cohen], and we denote it by $\mathsf{css}(K) \in H^4(K; \Theta)$.

This definition includes manifolds with boundary, and the invariant of the boundary is $\mathsf{css}(\partial K) = i^*\mathsf{css}(K)$, where $i^*: H^4(K) \to H^4(\partial K)$ is induced by inclusion. The key result is that the Rokhlin homomorphism relates the Kirby–Siebenmann and CSS invariants:

Proposition 3.3. *If $r: M \to K$ is a manifold resolution of a PL homology manifold, then $r^*(\mathsf{rok}(\mathsf{css}(K)) = \mathsf{ksm}(M)$.*

The usual formulation is for the special case with r a homeomorphism. The additional information in the resolution version is that including the vertex singularities makes no difference. They neither contribute additional problems, nor do they give a way to avoid any of these problems.

The description of css should make this result very plausible, and if the definition of the Kirby–Siebenmann invariant is understood (which we won't do here), the homeomorphism version should be obvious. The resolution version follows easily from the homeomorphism version and the uniqueness of resolutions up to homeomorphism.

3.8 The Homology Picture

The dual homology class is sometimes easier to work with but takes more care to define correctly. The basic idea is to use simplicial chains and represent the class in $Z[(n-4)\text{-simplices}] \otimes \Theta$ by using the class of the link of a simplex σ as the coefficient on σ. There is a problem with this: an orientation is required to define an element in Θ, but the data provides an orientation for the simplex rather than the dual cone. An orientation of the manifold can be used to transform simplex orientations to dual-cone orientations, but being too casual with this invites another mistake: the invariant is in twisted homology.

A homology manifold has a double cover with a canonical orientation, $\hat{K} \to K$. The group of covering transformations is $Z/2$ and the generator acts on $\hat{K}$ by interchanging sheets and (therefore) reversing orientation. Consider the simplicial chains $C^{\Delta}(\hat{K})$ as a free complex over the group ring $Z[Z/2]$, and suppose A is a $Z[Z/2]$ module. We define the homology $H_n(\hat{K}; A)$ to be the homology of the complex $C^{\Delta}(\hat{K}) \otimes A$, where the tensor product is taken over $Z[Z/2]$.

If $Z/2$ acts trivially on A, then the tensor product kills the action on the chains of $\hat{K}$ and the result is ordinary homology. We will be concerned with the opposite extreme, $A = Z$ with $Z/2$ acting by multiplication by -1.

After this preparation we can define the Cohen–Saito–Sullivan *homology* class by

$$\mathsf{css}_*(K) = \Sigma_\sigma \sigma \cdot [\mathsf{link}(\sigma)] \in H_{n-4}(\hat{K}, \partial\hat{K}; \Theta)$$

where $Z/2$ acts on Θ by reversing orientation, and the orientation of $\mathsf{link}(\sigma)$ is induced by the orientation of σ and the canonical orientation of $\hat{K}$.

This definition also includes manifolds with boundary, and the invariant of the boundary is given by the boundary homomorphism in the long exact sequence of the pair.

The homology and cohomology definitions are Poincaré dual. Duality between simplices and dual cones is particularly clear: each simplex intersects exactly one dual cone (its own) in a single point, and this pairing gives a chain isomorphism between simplicial homology and dual-cone cohomology when links are homology spheres.[4] This pairing matches up the two definitions.

4 Gauge Theory

The Casson invariant (see [am]) gave the first hint that something like gauge theory would play a role in this story. Casson used representation varieties and Heegard decompositions to define an integer-valued invariant of homology 3-spheres, and showed that the mod 2 reduction is the Rokhlin invariant. However it is a invariant of diffeomorphism type, not homology H-cobordism. It does not define a function $\Theta \to \mathbf{Z}$, and has little consequence for the triangulation questions.

Fintushel and Stern [fintstern90] used the Floer theory associated with Donaldson's anti-self-dual Yang–Mills theory to show that certain families of Seifert fibered homology 3-spheres are linearly independent in Θ. The families are infinite so Θ has infinite rank. This implies that most manifolds have vastly many concordance classes of triangulations, but does not clarify the existence question because all these homology 3-spheres are in the kernel of the Rokhlin homomorphism.

There has been quite a bit of work done since Fintushel–Stern, with invariants derived from gradings in various Floer homology theories; see Manolescu's discussion of Frøyshov correction terms. The next qualitatively new progress, however, is in Manolescu's paper. The outcome is three functions $\Theta \to \mathbf{Z}$ which are not homomorphisms, but have enough structure to show that a homology sphere with nontrivial Rokhlin cannot have order 2 in Θ. This implies that manifolds whose Kirby–Siebenmann classes do not lift to mod 4 cohomology, cannot be triangulated. See [gs5] for a 5-dimensional example. Somewhat more elaborate arguments with these functions seem to show that many Rokhlin-nontrivial spheres have infinite order. The full consequences are not yet known.

Sections 4.1–4.6 give a qualitative outline of the preprint version of Manolescu's paper. The published version may be different. References such as "[mano13, §3.1]" are abbreviated to "M3.1," and readers who want to see things like the Chern-Simons-Dirac functional written out should refer to this paper. Alternate perspectives for experts are suggested in Sects. 4.7–4.8

[4] This is, in fact, Poincaré's picture of duality, and will be discussed further in the history section.

4.1 Physics Description

The idea, on the physics level of rigor, is that the Floer homology theory associated with the Seiberg–Witten equations is given by the Chern-Simons-Dirac functional on an appropriate function space. This functional is invariant under a big symmetry group. Divide by the symmetry group, then we want to think of the induced function on the quotient as a sort of Morse function and study its gradient flow. More specifically, we are concerned with the finite-energy trajectories. The quotient is infinite-dimensional, but we can enclose the finite-energy trajectories in an essentially finite-dimensional box. Invariants of the system come from algebraic-topological invariants of this box.

This description offers an alarmingly large number of ways to misunderstand the construction, and one goal is to clarify the strategy and logical structure of the process. For instance, finite-dimensional differential and algebraic topology are mature subjects with a lot of sharp tools. It is useful to see the infinite-dimensional part of the analysis as a sequence of reductions designed to bring part of the structure within range of these sharp tools. Another, possibly dubious, goal is to try to clarify features of the technical details and how they might be sharpened, but without actually describing the details. Finally, the analysis described is for three-manifolds whose first homology is torsion ($b_1 = 0$). The analysis in the general case is considerably more elaborate.

4.2 The Coulomb Slice

The first step in the heuristic description is to "divide by an infinite group of symmetries." It is almost impossible to make literal sense of this, and in M3.1 Manolescu uses the Coulomb slice to avoid it. There is a ("normalized") subgroup of the full symmetry group with the property that each orbit intersects this slice in exactly one point. The slice is therefore a model for the quotient by this subgroup, and projection to the slice reduces the symmetry to the quotient of whole group by the subgroup. The quotient is the compact Lie group $Pin(2)$.

Since it is compact, dividing by $Pin(2)$ makes good sense, but it introduces singularities that are much more painful than symmetry groups. The plan is therefore to do a nonsingular equivariant reduction to finite dimensions, and the long-term strategy is roughly "let the finite-dimensional people deal with the group action."

Manolescu explicitly describes the restriction of the Chern-Simons-Dirac functional to the Coulomb slice, and describes a projected Riemannian metric that converts the derivative of the functional to a gradient vectorfield with the property that the projection preserves gradient flows. This description usually gives non-specialists the wrong picture because the "Riemannian metric" is not complete. The slice is a Frechét space of C^∞ functions, and there is no existence theorem for flows in this context. In fact, in most directions the gradient vectorfield does not have a flow, even for short time, and the flow trajectories exploited by Floer and others exist due to a regularity theorem for solutions of a differential equation with boundary conditions. In other words there is only a small and precious fragment of a flow for this vectorfield, and this is *not* Morse theory with a globally defined flow. The observation that projection to the Coulomb slice preserves flows means it preserves this small and precious fragment, not something global.

This explanation is still not quite right. Manolescu doesn't actually identify the flow fragment in infinite dimensions, so saying that the projection preserves whatever part of the flow that happens to exist is a heuristic summary. On a technical level the projection preserves *reasons* the fragment exists and it is these reasons, not the flow itself, that power the rest of the argument.

4.3 Sobolev Completions

In the last paragraph of M3.1 the space $V_{(k)}$ is defined as the completion of the Coulomb slice, using the L^2 Sobolev norm on the first k derivatives. This gives Hilbert spaces but still doesn't give us a flow because the "vectorfield" now changes spaces: it is of the form

$$\ell + c\colon V_{(k+1)} \to V_{(k)}$$

with ℓ linear Fredholm and c compact.[5] The index shift corresponds to a loss of a derivative, reflecting the fact that we are working with a differential equation. The maneuvering (bootstrapping) needed to more-or-less recover this lost derivative is a crucial analytic ingredient. Almost nothing is said about this in [mano13], but some details are in [mano03, Sect. 3 and 4], phrased in terms of flows rather than vectorfields. Manolescu's next step is projection to finite-dimensional spaces where there are well-behaved flows. There would be significant advantages to connecting directly with Morse theory in an infinite-dimensional setting rather than in projections; see section "Hilbert, or SC Manifolds" for further comments.

4.4 Eigenspace Projections

In section M3.2 Manolescu defines V_τ^ν to be the subspace of V spanned by eigenvectors of ℓ with eigenvalues in the interval $(\tau, \nu]$. This uses the fact, prominent in [mano03] but unmentioned in [mano13], that ℓ is self-adjoint. In particular its eigenvalues are real and eigenspaces are spanned by eigenvectors. These spaces are

[5] The spaces V are vector spaces and ℓ linear because we are assuming $b_1 = 0$ (homology sphere). The general situation is more complicated.

finite-dimensional because ℓ is Fredholm. Finally, the symmetry group $Pin(2)$ acts on them because they are defined using equivariant data.

There is a technical modification that deserves comment. The orthogonal projections $V \to V_\tau^\nu$ give a function from the parameter space $\{\tau < \nu\}$ to linear maps $V \to V$. This takes discrete values (depending only on the eigenvalues in the interval) so is highly discontinuous. Manolescu smooths this function: the dimension of the image still jumps but when it does, the projection on the new part is multiplied by a very small number. The result is continuous as a function into the space of linear maps. This implies that the finite projections of the CSD vectorfield become smooth functions of the eigenvalue parameters. This is useful in showing that parts of the qualitative structure of the output do not depend on the parameters once they are sufficiently large.

The final modification of the flow is done in section M3.7. There is a unique reducible solution of the equations, and non-free points of the $Pin(2)$ action come from this. The functional is perturbed slightly (following the earlier [mano03]) to make the reducible solution a nondegenerate critical point. The irreducible critical points can also be made nondegenerate in an appropriate equivariant (Bott) sense.

This is one of the places where the Seiberg–Witten theory diverges in a qualitative way from the Donaldson theory. The finite-energy trajectories in the Donaldson–Floer theory cannot be made nondegenerate, and the analysis takes place on a center manifold. This is rather more delicate.

4.5 Isolated Invariant Sets

The last structural input from the infinite-dimensional context is specification of the "precious fragment" of the flow supposed to have come from infinite dimensions. This is done by Proposition M3.1 in [mano13], which is a reference back to Proposition 3 of [mano03]. The flow fragment is the union of trajectories that stay in a ball of a certain radius, and the key fact is that it is isolated in the sense that it is the same as the union of trajectories that stay in a ball of twice the radius.

We comment on the logic of the reduction. Defining the invariant uses only the answer (the form of the explicit finite-dimensional approximations) and the proof in Proposition M3.1 that the trajectories-in-a-ball construction gives an isolated invariant set. This does not use the construction of a flow fragment in infinite dimensions, so the demonstration that such a flow fragment would have been preserved by the projection is not actually used. This demonstration does, however, give a tight connection between this construction and those of Floer et.al. that do use the infinite-dimensional flow.

4.6 Equivariant Stable Homotopy Theory

The plan is to enclose the isolated invariant set identified in the previous step in a nice box, and extract information about the system from algebraic and geometric topology of the box. The box is a subspace (or submanifold) of a finite-dimensional vector space so this is the point at which the problem enters the finite-dimensional world. Manolescu is not a native of this world, however, and his treatment could be refined. We briefly sketch Manolescu's definition of the invariants in this section. The main difficulties come in showing that these are well-defined and have good properties. The next section hints at some of these difficulties and suggests approaches that may be better adapted.

Manolescu uses the Conley Index construction to get a "box" enclosing the isolated invariant set. The output is a pair of spaces that depends on choices or, by taking the quotient, a pointed space that the choices change only by homotopy equivalence. This takes place in an eigenvalue projection V_τ^ν and changing τ and ν changes the pointed-space output by suspension. The object associated with the system is therefore a spectrum in the homotopy-theory sense. Finally, all these things have $Pin(2)$ actions, and the suspensions are by $Pin(2)$ representations. The proper setting for all this is evidently some sort of equivariant stable homotopy theory. The most coherent account in the literature is Lewis–May–Steinberger [may86], and Manolescu uses this version. The next step is to extract numerical invariants from these $Pin(2)$-equivariant spectra using Borel homology.

To a first approximation the homology appropriate to a G-space X is the homology of the quotient X/G. This works well for free actions but undervalues fixed sets. The Borel remedy is to make the action free by product with a contractible free G-space EG, and take the homology of the quotient $(X \times EG)/G$. The free part of X is unchanged by this but points fixed by a subgroup $H \subset G$ are blown up to copies of the classifying space EG/H. These classifying spaces are usually homologically infinite-dimensional, so fixed sets become quite prominent. Another benefit of the Borel construction is that the homology of $(X \times EG)/G$ is a module over the cohomology of $BG := EG/G$. These facts are illustrated by a localization theorem quoted in M2.1: suppose X is a finite G-complex[6] and the action is free on the complement of $A \subset X$. A localization that kills finite-dimensional $H^*(BG)$ modules kills the relative Borel homology of (X, A), so the inclusion $A \to X$ induces an isomorphism on localizations. There is a difficulty that Borel homology is not fully invariant under equivariant suspensions. Manolescu finesses this with F_2 coefficients, but eventually it must be instituted.

In the case at hand the G-objects are spectra rather than single spaces. X can be thought of as the equivariant suspension spectrum of a finite G-complex and the sub-spectrum A of non-free points is essentially the suspension spectrum of a point. Inclusion therefore gives a $H^*(BG)$-module homomorphism $H_*(BG) \to$

[6]The finiteness hypothesis on X is missing in the statement in [mano13].

$H_*((X \times EG)/G)$. Finiteness of X implies that the third term in the long exact sequence (the homology of the free pair (X, A)) is finite-dimensional. In particular the kernel of $H_*(BG) \to H_*((X \times EG)/G)$ is a finite-dimensional $H^*(BG)$-module. When $G = Pin(2)$ these submodules are characterized by three integers α, β, γ, and these are Manolescu's invariants. The algebraic details give a pretty picture, and readers should refer to Manolescu's paper for this.

4.7 Handcrafted Contexts

Both stable homotopy theory and equivariant topology are sprawling, complicated subjects. Off-the-shelf versions tend to be optimized for particular applications and often use shortcuts or sloppy constructions that can cause trouble in other circumstances. The best practice is to handcraft a theory that fits the application, but this requires insider expertise. In this section we suggest such a handcrafted context for the finite-dimensional part of Manolescu's development.

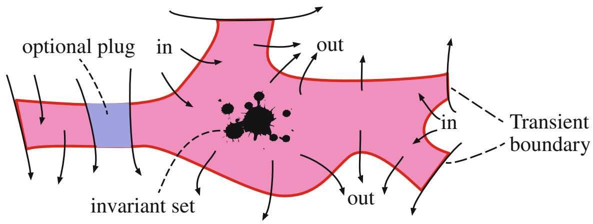
Lyapunov Blocks

The first step is to be more precise about the data at the transition from analysis to finite-dimensional topology. Manolescu uses the Conley index construction to get a "box" enclosing an isolated invariant set in a flow on a manifold V. We recommend instead an object we call a *Lyapunov block*. These were introduced and shown to exist using Lyapunov functions by Wilson and Yorke [wilsonyorke73], and shown to be essentially equivalent to Lyapunov functions by Wilson [wilson80]. Wilson and Yorke call these "isolating blocks," but a more distinctive name seems to be needed. This construction has been revisited recently by Cornea [cornea], Rot–Vandervorst [rotvdh], and others.

A Lyapunov block for an invariant set in a flow is a compact smooth codimension-0 submanifold-with-corners $B \subset V$ with boundary divided into submanifolds $\partial_- B \cup \partial_0 B \cup \partial_+ B$. Trajectories intersect B in arcs. Trajectories enter through the incoming boundary $\partial_+ B$, exit through the outgoing boundary $\partial_- B$. The transient boundary $\partial_0 B$ is a union of intersections with trajectories, and the trajectory arcs give a product structure $\partial_0 B \simeq \partial_{0,+} B \times I$; see Fig. 2. Finally, the trajectories completely contained in B are those in the original invariant set. The underlying smooth manifold structure can be thought of as a smooth manifold triad $(B, \partial_+ B, \partial_- B)$. $\partial_0 B$ is a collar so absorbing it into either $\partial_- B$ or $\partial_+ B$ (or half into each) changes them only by canonical diffeomorphism.

These blocks are not well-defined: different choices in a truncation step give B that differ by addition or deletion of plugs of the form $P \times I$, that intersects trajectories in product arcs $\{p\} \times I$. This implies that the pairs $(B, \partial_+ B)$ and $(B, \partial_- B)$

Fig. 2 A Lyapunov block for a flow



have well-defined *relative* homotopy types.[7] To relate this to Manolescu's version, $(B, \partial_{-}B)$ is a particularly nice Conley index pair for the flow, and the index itself is the pointed space $B/\partial_{-}B$. The quotient $B/\partial_{+}B$ is a Conley index for the reversed flow. The manifold triad therefore gives both Conley indices and precisely encodes their relationship.

Smooth Manifold Triads

The handcrafted context appropriate to this situation seems to be a stable category defined using equivariant smooth manifold triads. It is "stable" in the sense that the objects are families of triads related by equivariant suspensions that are "internal" in the sense that they come from eigenvalue-range changes (see below). This context receives Lyapunov blocks without further processing. It has many other virtues, as we explain next, and in fact we like Lyapunov blocks because they permit use of this context.

This context does *not* follow the standard practice of dividing to get pointed spaces. Data from geometric situations often comes as pairs with structure that does not gracefully extend to pointed-space quotients. Bundles on pairs, for instance, rarely extend to the pointed space. This means they have to be described as "bundles over the complement of the basepoint," and to work with them one must recover the pair by deleting a neighborhood of the basepoint. Group actions can be extended to have the quotient basepoint as a fixed point, but this is often just cosmetic. In many geometric applications, for instance, algebraic topology is done equivariantly on the universal cover. The fact that the action is free is essential. The pointed-space quotient therefore must be described as an action free in the complement of the basepoint, and again much of the work requires deleting the basepoint to recover a pair with a free action. Having to delete the basepoint is a clue that dividing to get a basepoint was a mistake. In some cases the pair information can be recovered stably without explicitly deleting the basepoint, but it is usually a lot of work.

The second advantage of this context is that in the manifold-triad world, Spanier–Whitehead duality is implemented by interchanging the two boundary components.

[7] Blocks can be modified to eliminate the transient boundary [rotvdh], but it is best not to make this part of the definition because it makes the "plug" variation hard to formulate.

Interchanging boundaries in a Lyapunov block corresponds to reversing the flow, so it is obvious that the flow and its reverse have S-W dual blocks. In the pointed-space context there is a stable description in terms of maps from a smash product to a sphere, but this is a characterization, not the definition, and it does not work in all cases. Manolescu quotes this version in M2.4, and cites references that show that the stable and unstable Conley indices are S-W dual in this sense. But these references use Lyapunov blocks, so the net effect is "discard the Conley constructions and redo the whole thing with manifold triads." Going back to Conley indices not only is inefficient but also introduces troublesome ambiguities about suspensions. This difficulty is discussed next.

The final wrinkle in this context has to do with the meaning of "suspension." Enlarging the range of eigenvalues changes the projection by product with a representation of $Pin(2)$, and changes the Lyapunov block by suspension with the ball in this representation. This is an "internal" suspension because it is specified by the analytic data. Understanding how internal suspensions change, for instance, when the metric on the original manifold is varied, is a job for analysis. External suspensions used to define equivariant invariants are specified differently, and the two types of suspensions should be kept separate. In particular the eigenvalue-change suspensions should not be seen as instances of external suspension operations. To explain this, note that the equivariant theory of Lewis et al. [may86] (used by Manolescu) is handcrafted to give a setting for homology theories and classifying spaces. Roughly speaking, they want to grade homology theories by equivalence classes of objects in the category of representations. When objects have nontrivial automorphisms, equivalence classes of objects do not form objects in a useful category. The standard fix for this is to use a skeleton subcategory with one object in each equivalence class. In the equivariant setting this means choosing one representation in each equivalence class, and always suspending by exactly *this* representation. This is fine for *external* suspensions, but representations that come internally from eigenvalue projections have no canonical way to be identified with randomly chosen representatives. If the group is S^1, as in most previous work on Seiberg–Witten–Floer theory, then there are essentially no automorphisms and this issue can be finessed. Manolescu's key insight, however, is that $Pin(2)$ is the right symmetry for this problem,[8] and these representations have automorphisms that make identifications problematic. The solution is to avoid using external suspensions in describing the geometric invariant. Lyapunov blocks do this.

4.8 Next Goals

Floer homology is, to a degree, a solution in search of worthy problems. Distinguishing knots is a baby problem whose persistence just reflects the lack of real work to do. The triangulation problem is useful for teething technology but, as explained in

[8] The $Pin(2)$ symmetry was observed much earlier, cf., [bauerFuruta], but not fully exploited.

the History section, is a backwater with no important application. This weakness is reflected in the structure of Manolescu's invariant: detailed information about homology spheres lies in the part of the moduli space on which $Pin(2)$ acts freely, but the invariant discards all this except the levels at which it cancels homology coming from the fixed point. This is not a gateway to something deeper. We have several suggestions for further work.

Complicated Geometric Structure

The first suggestion is motivated by internal structure of the analytic arguments. Analysis associates to a homology 3-sphere a complicated $Pin(2)$-equivariant gadget. We expect this to reveal something about geometric properties of the three-manifold, but neither the properties nor the mechanisms of revelation are clear. A useful intermediate step would be a complicated $Pin(2)$-equivariant gadget derived more directly from the topological object. The two equivariant gadgets might be related by a sort of index theorem. The point is that sometimes it is easier to relate two complicated things than to understand either in detail, and the connection can be a powerful aid to understand. We suggest maps from the homology sphere to $S^3 \simeq SU(2)$ as the topological gadget. $Pin(2)$ acts on this because it is a subgroup of $SU(2)$, and these maps should connect to geometric structure by a form of generalized Morse theory, cf., [gaykirby].

Hilbert, or SC Manifolds

The key analytic goal is to situate the objects of interest in a context accessible to "finite-dimensional" geometric and algebraic topology. The context does not have to be literally finite-dimensional to use the techniques, however, and a context that does not require finite-dimensional projections would simplify formulation of invariants. The first requirement for such a context is an effective global existence theorem for flows. There seem to be at two possibilities that are, in a sense, at opposite extremes.

Manolescu begins (see Sect. 4.2) by restricting the Chern-Simons-Dirac functional to the Coulomb slice, and using a Riemannian metric to convert the derivative of the functional to a vector field. It would be quite natural to complete with respect to this metric, to get a vectorfield on a separable Hilbert manifold. The problem is that current estimates are not good enough to show that the finite-energy trajectories form an isolated invariant set in this topology. There are heuristic reasons to worry that they are not isolated in general. A perturbation of the system to be "nondegenerate" in some sense might help. Eventually the geometric invariant would be a Lyapunov block in the Hilbert manifold, together with an equivalence class of *structures* related to the finite-dimensional projections. This would clarify that the objects obtained by projection are fragments of a structure on the invariant object, not the invariant object itself. A Hilbert-manifold formulation should be much easier to extend to things like Hilbert-manifold bundles over $H_1(X; R)$, which may be necessary for three-manifolds with $\beta_1 \neq 0$.

Another possible context is the Banach-scale manifolds developed by Hofer, Wysocki, and Zehnder. The Hilbert approach takes place at a fixed level of differentiability, while the Banach-scale approach organizes the way in which function spaces of increasing differentiability approach C^∞. It is "handcrafted" in the sense of Sect. 4.7 to formalize and exploit the bootstrapping common in applications.

The first comment is that the Hofer–Wysocki–Zehnder "polyfold" theory is not appropriate here. This was developed to handle closure problems in quotients. Here this is handled by taking the quotient by a subgroup of the full gauge group which, since it has a global slice, has no closure problems. This leaves a residual gauge action by $Pin(2)$. Dividing by this does introduce orbit-closure problems but (1) these seem to be outside the reach of the Hofer–Wysocki–Zehnder polyfold theory, and (2) by now it should be quite clear that equivariant nonsingular objects are more effective than trying to describe some sort of structure on singular quotients. The second comment is that a useful version of "Lyapunov block" would be needed, and this may require negotiation with the topological theory that has to use it. The final comment is that this may give a setting for the germ-near-a-compact-set suggestion in the next section.

Stay in Dimension 4

The motivation for the final suggestion is external to the analysis. The best guides to development of a theory are deep potential applications. Floer homology of three-manifolds is supposed to organize boundary values and glueing properties of gauge theories on smooth four-manifolds but, in general, three-manifolds slices and boundaries do not adequately reflect the complexity of smooth four-manifolds. We explain this in a context that ideally would connect with homology 3-spheres.

Suppose M is a smooth four-manifold with a submanifold V *homeomorphic* to $S^3 \times \mathbf{R}$. If M is compact, simply connected, and V separates M, then a relatively soft argument [freedmanTaylor] shows that M also contains a smooth homology 3-sphere homologous to $S^3 \times \{0\}$. But this is usually not true if M is either noncompact or not simply connected. For instance, a compact four-manifold has a smooth structure in the complement of a point, and this point has a neighborhood homeomorphic to $S^3 \times \mathbf{R}$, but almost none of these contain smooth homology spheres. When there is an appropriate homology sphere in M it is usually not in the given V.

Another soft argument shows that in the compact simply connected case any two homology 3-spheres arising as above are homology H-cobordant, but not "in M". Note that disjoint homologous homology spheres have a region between them that is a H-cobordism. Ideally, if we have two homology spheres, then we would find a third homology sphere disjoint from both. The first two would both be H-cobordism to the third, so the first two would be H-cobordant by a composition of *embedded*

H-cobordisms. Unfortunately we can usually not find a third disjoint sphere, and the soft argument does not give embedded H-cobordisms.

The moral of this story is that we can use transversality to get smooth three-manifold splittings, but these three-manifolds usually cannot reflect the global homotopy theory of the manifold even up to homology. A glueing theory that depends on finding nice slices (e.g., smooth homology spheres in topological connected sums) therefore cannot be an effective general setting.

A better setting for glueing 4-d theories should be some sort of "germs of necks" that locally separate the four-manifold. We have much better criteria for finding good *topological* slices in four-manifolds, so a first approximation would be "germs near $X \times \{0\}$ of smooth structures on $X \times \mathbf{R}$," where X is a closed three-manifold, but the smooth structure on $X \times \mathbf{R}$ is not the product structure.

Smooth neighborhoods of topological embeddings are the sort of mixed-category thing that (according to the History section) is probably a bad idea in the long term, but it gives a concise starting point. The homotopy data required to find a topological slice in a "neck" are non-obvious and fairly elaborate. The data needed to find a "virtual analytic slice" may also be elaborate, so speculations should wait on feedback from analysis. In any case the point for the present discussion is that the best next step in Floer-type theory is probably gauge theory on 4-d "neck germs," not gauge theory on three-manifolds.

5 History

Poincaré's insights about the homology of manifolds, at the end of the nineteenth century, are usually celebrated as the starting point of modern topology. But many of his insights were wrong in detail, and his methodology was so deficient that it could not be used as a foundation for further development. His contemporaries found it inconceivable that the Emperor might have no clothes, so they spent the next quarter-century trying to see them. Kneser's triangulation questions are precise formulations of what it would take to make Poincaré's arguments sensible. Labeling one of them "the Hauptvermutung" suggests that he still hoped it would all work out. But it did not. As interesting as these questions seem, they are a technical dead end. Not only are they not a foundation for manifold theory, but they seem not to have significant applications. Details of this story, and how topology finally recovered from Poincaré's influence, are told in this section.

5.1 Pre-modern Methodology

Poincaré worked during the period when modern infinite-precision mathematics was being developed [rev]. He was not part of this development, however, but worked in—and strongly defended—the older heuristic and intuitive style.

His explanations often included the technical keys needed for a modern proof of a modern interpretation of his assertion. But he often omitted hypotheses necessary for his assertions to be correct, and his arguments were too casual to reveal the need for these hypotheses. He gave examples, but did not use precise definitions and often did not verify that the examples satisfied the properties he ascribed to them. This casual approach, and the philosophical convictions that underlay it, made for a difficult start for the subject.

For instance, Poincaré proceeded on the presumption that the choice of analytic, combinatorial, or topological tools would be dictated by the task at hand rather than the type of object. Functionally this amounts to an implicit claim that topological, PL, and smooth manifolds are all the same. Clearly anything built on this foundation was doomed. But identifying this as a flaw in Poincaré's work would have invited strong political and philosophical attack and the new methodologies were not secure enough for this. Kneser's triangulation questions 25 years were precise technical formulations of what would be needed to justify Poincaré's work, but he still did not identify this as a gap in the work.

Not only was it hard to know which parts of Poincaré's work were solid, but also apparently it was hard to track which parts were actually known to be false. For instance, in a 1912 paper of Veblen and Alexander [veblenAlex12] we find

> Poincaré has proved that any manifold M_n may be completely characterized from a topological point of view by means of suitably chosen matrices ...

This refers to the 1895 claim that the incidence matrices of a triangulation (now called boundary homomorphisms in the chain complex) characterized manifolds up to homeomorphism. We overlook this blunder today because Poincaré himself disproved it not long after, by using the fundamental group to show the "Poincaré sphere" is not S^3 even though it has equivalent chains. But more than 10 years later Veblen and Alexander seem to have been unaware of this refutation.

5.2 Poincaré's Duality

An explicit example of Poincaré's methodology is provided by his description of duality. He observed the beautiful pairing of simplices and dual cones explained in Sect. 3.8. But he called these dual cones "cells," and implicitly presumed that they were equivalent (in an unspecified sense) to disks. Instead of seeing this as a general PL construction that might or might not give a cell, it was seen as a manifold construction that "failed" if the output was not a cell. This convention makes arguments with dual cells logically sensible, but it hides the necessity of showing that the construction does not fail in specific instances. One of Poincaré's classes of examples was inverse images of regular values of smooth maps $R^n \to R^k$. In what sense can these be triangulated, and why are the dual objects cells? Whitehead sorted this out some 40 years later [whitehead40]. The proof was probably beyond Poincaré's ability, but the real problem was that he did not notice (or acknowledge) that there was a gap.

Another difficulty is that Poincaré's duality relates different objects: homology (or Betti numbers) based on simplices, on one hand, and homology based on dual cells, on the other hand. In order to get duality as a symmetry of a single object, these must be identified in some other way. This time Poincaré got it wrong: dual cells actually give cohomology so, as we know now, duality gives an isomorphism between homology and cohomology. The homology/cohomology distinction (in the group formulation) together with the Universal Coefficient Theorem explain why the torsion has symmetry shifted one dimension from the Betti-number symmetry. Poincaré missed this, and found a patch only after Heegard pointed out a contradiction. In another direction, duality requires some sort of orientation and (as we saw with Θ in Sect. 3.8) may be twisted even when there is an orientation. When the manifold has boundary, or is not compact, duality pairs homology with rel-boundary or compact-support homology. Homology of the boundary appears as an error term for full symmetry. Again these results were beyond Poincaré's intuitive definitions and heuristic arguments, but the real problem was that he did not notice (or acknowledge) that more precision was needed.

5.3 Point-Set Topology

Schoenfliss and others were developing point-set topology around the same time, and the relationship between the two efforts is instructive.

An important point-set goal was to settle the status of the Jordan Curve theorem. This is not hard to prove for smooth or PL curves, but an intuitive extrapolation to continuous curves was discredited by the discovery of continuous space-filling curves by Peano and others. The continuous version had important implications for the emerging role of topology as a setting for analysis. For instance, integration along a closed curve around a "hole" in the plane was a vital tool in complex analysis. Integration required piecewise-smooth curves. The question was: were "analytic holes" identified by piecewise smooth curves the same as "topological holes" identified with continuous curves? If not then the role of general topology would probably be quite limited.

Addressing the Jordan Curve problem turned out to be difficult, and fixing gaps in attempted proofs required quite a bit of precision about open sets, topologies, separation properties, etc. In short, it required modern infinite-precision techniques. Wilder [hist] found it curious that Shoenfliss never mentioned Poincaré or his work, since nowadays the Jordan Curve theorem and high-dimensional analogues are seen as immediate consequences of a homological duality theorem. But this makes sense: Shoenfliss was trying to fix a problem in a heuristic argument, and Poincaré used heuristic arguments. The duality approach was not available to Shoenfliss because—for good reason—he could not trust Poincaré's statements about duality.

5.4 Constrained by Philosophy

The general question in this section is: why did it take Poincaré's successors so long to find their way past his confusions? The short answer is that they were in the very early part of the modern period and still vulnerable to old and counterproductive convictions. We go through some of the details for what they reveal about the short answer: what *were* the counterproductive nineteenth-century convictions, and how did they inhibit mathematical development?

To be more specific, a mathematician with modern training would probably respond to Poincaré's work with something like

> The setting seems to be polyhedra, and the key property seems to be that the dual of a simplex should be a cell. Let's take this as the working definition of 'manifold', and see where it takes us. Later we may see something better, but this is a way to get started.

We now know that the basic theory of PL manifolds is more elementary and accessible than either smooth or topological manifolds, and this working definition is a pretty good pointer to the theory. Why were Poincaré's successors slow to approach the subject this way, and when they did, why did it not work as well as we might have expected?

The first problem was that Poincaré and other nineteenth-century mathematicians objected to the use of explicit definitions. The objection goes back 2400 years to Pythagoras and Plato, and is roughly that accepting a definition is like accepting a religious doctrine: you get locked in and blocked from any direct (intuitive) connection to "reality." The precise-definition movement reflects practice in science: established definitions are distillations of the discoveries of our predecessors, and working definitions provide precise input needed for high-precision reasoning. It is odd that this aspect of scientific practice came so late to mathematics, but recall that in the nineteenth century there was still a strong linkage between mathematics and philosophy. And still to this day, accepting a definition in philosophy is like accepting a religious doctrine.

An interesting transitional form appears in a long essay by Tietze in 1908 ([tietze]; see the translation at [tietzeTrans]). He defined manifolds as polyhedra such that the link of a simplex is simply connected, but did not define "simply connected." It is hard to imagine that he meant this literally. The use of the terminology "simply connected" indicates familiarity with Poincaré's work with the fundamental group, but Poincaré asks explicitly if it is possible for a three-manifold "to be simply connected and yet not a sphere." Simply connected is obviously wrong one dimension higher. His use of the term seems to have been a deliberately ambiguous placeholder in a proposal for a "big-picture" view of manifolds. This reflects the philosophical idea that big pictures should be independent of details, and the goal of heuristic arguments in the nineteenth-century tradition was to convince people that this was the right intuition, not actually prove things. On a practical level, Tieze may have been mindful of the advantages ambiguity had for Poincaré. People worked hard trying to find interpretations of Poincaré's ideas that would make them correct, but could not have been so generous if he had tried to be more precise and guessed wrong.

The next milestone we mention is the introduction of PL homology manifolds as a precise setting for the study of duality. Wilder [wilder] attributes this to Veblen in 1916. They knew these were not always locally Euclidean so would not be the final context for geometric work, but they would serve for algebraic topology until the geometric people got their acts together.

In the geometric line at that time, people were experimenting with various precise replacements for Tieze's placeholder. The favorite was "stars homeomorphic to Euclidean space." Today we would see this as a mixed-category idea that for general reasons is unlikely to be correct and in any case is inappropriate for a basic definition. This experience was not available at the time, of course, but they were not having success with homeomorphisms and there were clues that an all-PL version would have advantages. Why did they stick with homeomorphisms for so long? There were two philosophical concerns and a technical problem.

The first philosophical concern was that a "manifold" should be a *thing*. A topological space was considered a primitive thing,[9] and a space that satisfies a property (e.g., locally homeomorphic to Euclidean space) is a thing. A simplicial complex is also a thing. A polyhedron, however, is a space with an equivalence class of triangulations. This is a *structure* on a thing, not a primitive thing, so for philosophical reasons could not qualify as a correct definition of "manifold." This objection also blocked the use of coordinate charts to globalize differential structures.

The second philosophical objection to PL manifolds has to do with the "recognition problem." A simplicial complex is a finite set of data. Suppose someone sent you one in the mail. How would you know whether or not links of simplices were PL equivalent to spheres? Suppose the sender enclosed a note asserting that this was so. How could you check to be sure it was true? Bertrand Russell summarized the philosophical attitude toward such things [russell, p. 71]

> The method of "postulating" what we want has many advantages; they are the same as the advantages of theft over honest toil.

The manly thing to do, then, is to *prove* links are PL spheres, and *assuming* this is cowardly and philosophically dishonest. Today we might wonder that assuming that a space is locally Euclidean (rather than recognizing it as being so) is ok, while assuming PL is not. At any rate one consequence was that the generalized Poincaré conjecture[10] (then referred to as "the sphere problem") seemed to be essential to justify work in higher-dimensional PL manifolds. The effect was to paralyze the field.

---

[9] We now think of a topological space as a structure (a topology) on a set. In the Poincaré tradition, spaces were primitive objects with properties extrapolated from those of subsets of Euclidean spaces.

[10] The generalized Poincaré conjecture is the assertion that a polyhedron that is known to be a PL manifold and that has the homotopy type of the sphere is PL equivalent to the sphere.

The technical problem had to do with the definition of "PL equivalence" of simplicial complexes. The modern definition is that they should have a common subdivision. This is very convenient technically because if you show some invariant does not change under a single subdivision then it must be a PL invariant. For traditionalists, however, it seemed uncomfortably existential. Equivalence of smooth or topological objects uses a nice concrete function with specific local properties; shouldn't PL follow this pattern? Brouwer, the great intuitionist, intuited a direct simplicial criterion for stars in simplicial complexes to be "Euclidean" and proposed this as a replacement for Tietze's placeholder. His intuition was ineffective, however, and later shown to be wrong.[11]

We finally come to Kneser's triangulation questions. In 1924 he gave precise formulations of what would have to be done to show that "polyhedron locally homeomorphic with Euclidean space" really did give a theory as envisioned by Poincaré, Tietze, et al. Whether he intended it or not, one message was roughly "enough sterile big-picture speculation; time to focus on what it would take to make it work." In particular, since Poincaré's use of dual cells gives duality between homologies defined with two different triangulations, the uniqueness of triangulations was needed to show Poincaré's claims about duality were correct. It must have seemed scandalous that this was still unresolved a quarter-century after Poincaré made the claims. We might also see Hilbert's influence in the concise straight-to-the-point formulations.

When Van der Vaerden surveyed manifold theory in 1928 he described it as a "battlefield of techniques." There had been advances in methodology but still no effective definitions and big issues were still unsettled. In fact the situation was already improving. In 1926 Newman [newman26] had published a version in which stars were still assumed homeomorphic to Euclidean space, but with complicated combinatorial conditions. This still didn't work, but in 1928 he published a revision [newman28] in which this was replaced by the common-subdivision version still used in the mature theory. PL topology was finally launched but, as it turned out, a bit too late.

5.5 *Overtaken*

Manifolds were supposed to be a setting for global questions in analysis, so smooth manifolds were the main goal. We have been following the PL topology developed to make sense of Poincaré's combinatorial ideas, but there are two other approaches that would have done this. The most effective is singular homology. This requires some algebraic machinery, but it is simpler than the PL development, much more

[11] In 1941, after the dust had settled, J. H. C. Whitehead reviewed the various proposals from the 1920s. Brouwer's proposal was particularly dysfunctional, and one has to wonder if he had actually tried to work with it in any serious way.

general, and connects better with analytic use of sheaves, currents, and deRham cohomology. The second is less effective but closer to Poincaré's ideas: show that smooth manifolds have standard (piecewise-smooth) triangulations. Remember that Kneser called the uniqueness-of-triangulations question "the Hauptvermutung" (principal assertion) because it would show that simplicial homology is independent of the triangulation. Either of the other approaches would have accomplished this, and therefore achieved the principal motivation of the PL development. The historical question should be: given the obvious importance of the questions, why did it take so long to find *any* of these solutions[12]? Slow development in the PL track is only interesting because the others were slow as well.

Singular homology probably developed slowly because it is so far outside the received wisdom from Poincaré. It requires algebraic apparatus and while we now see plenty of clues about this in Poincaré's work, these only became visible after Noether's promotion of abstract algebra as a context for such things. Čech's open-cover approach to homology also pushed things in this direction, but again this was outside Poincaré's vision.

The lack of an effective definition held up development of smooth manifolds, just as it held up development of PL. And, like PL and unlike the singular theory, no huge technical leaps were necessary: the main obstructions were ineffective intuitions and philosophical objections to structures defined with coordinate charts. These were finally overcome by Veben and Whitehead [vebWhitehead32] in 1932, and we can identify two things that made the advance possible. The first was a change to a more modern style that better reflects mathematical structure. Veblen and Whitehead did not give a philosophical argument or a speculative "big picture"; they developed enough basic structure (with technical details) to demonstrate conclusively that this was an effective setting for differential geometry. The second change was in the mathematical community. Young people were attracted by the power and depth of precise definitions and full-precision reasoning, and were more than ready to trade philosophy for success, while the old people committed to philosophy were fading away. These changes led to a great flowering of the differential theory, and it was the setting for some of the deepest and most remarkable discoveries of the second half of the twentieth century.

One consequence of the smooth-manifold flowering, and the development of singular homology, was a near abandonment of PL topology for several decades. It continued to be used in low dimensions due to low-dimensional simplifications (homology identifies two-manifolds). Enough of a community had been established to sustain some general activity, but it lacked the guidance of an important goal.

The 1950s and 1960s saw a renascence in PL topology. Smale's development of handlebody theory, and particularly his proof of a form of the generalized Poincaré conjecture, electrified the manifold communities. Smale's work was in the smooth world, coming from a study of the dynamics of Morse functions, but "handles" appear much more easily and naturally in PL. Milnor's discovery of

[12]Existence of piecewise-smooth triangulations was shown in 1940 by Whitehead [whitehead40].

multiple smooth structures on the 7-sphere [milnor56] was a huge boost. The reason was that Smale had proved that high-dimensional *smooth* homotopy spheres were *homeomorphic* to the sphere. Stallings then used PL techniques to show that high-dimensional PL homotopy spheres were homeomorphic to the sphere. Both of these had the defect that the conclusions were in a different category from the hypotheses. Smale improved on this by using PL versions of his techniques to show that high-dimensional PL homotopy spheres are PL isomorphic to spheres. Milnor's discovery showed that this is false in the smooth world, so PL is genuinely simpler and closer to the original intuitions.[13] All this took place in the modern links-are-PL-spheres context. Kneser's questions played no role and, as far as the main-line developments were concerned, were a dead-end curiosity.

By the end of the 1960s PL was again overshadowed, this time by development of purely topological manifold theory. Basic topological techniques are much more complicated than PL, almost insanely so in some cases, but the outcomes are more systematic and coherent. Further progress on what seemed to be PL issues also required outside techniques: the 3-dimensional Poincaré conjecture was settled by Perelman with delicate analytic arguments almost 80 years after Kneser's work, and 100 years after Poincaré hinted that this might be the key to further progress. The 4-dimensional case is still open in 2014, and no resolution is in sight. Finally, as we have seen here, insight into the structure of homology 3-spheres seems to require gauge theory.

5.6 Summary

Kneser's triangulation questions were a careful formulation of what it would take to develop a theory of manifolds that followed Poincaré's intuitions and nineteenth-century philosophy. Not long after, more fruitful approaches emerged based on full-precision twentieth-century methodology. Kneser's questions proved to be a curiosity: a nice challenge for developing technology, but apparently without significant implications.

Acknowledgements This work was partially supported by the Max Planck Institute for Mathematics, in Bonn.

[13] Milnor's discovery also invalidated the intuition, inherited from Poincaré, that there would be a single world of "manifolds" where all techniques would be available. Subsequent developments, as we have seen here, revealed how confining that intuition had been.

References

[am] S. Akbulut, J.D. McCarthy, *Casson's Invariant for Oriented Homology 3-Spheres: An Exposition*. Mathematical Notes, vol. 36 (Princeton University Press, Princeton, 1990)

[bauerFuruta] S. Bauer, M. Furuta, A stable cohomotopy refinement of Seiberg-Witten invariants. I. Invent. Math. **155**(1), 1–19 (2004)

[BFMW] J. Bryant, S. Ferry, W. Mio, S. Weinberger, Topology of homology manifolds. Ann. Math. (2) **143**(3), 435–467 (1996)

[cohen] M.M. Cohen, Homeomorphisms between homotopy manifolds and their resolutions. Invent. Math. **10**, 239–250 (1970)

[cornea] O. Cornea, Homotopical dynamics: suspension and duality. Ergod. Theory Dyn. Syst. **20**(2), 379–391 (2000)

[daverman] R.J. Daverman, *Decompositions of Manifolds* (AMS Chelsea Publishing, Providence, 2007). Reprint of the 1986 original

[davermanwalsh] R.J. Daverman, J.J. Walsh, A ghastly generalized n-manifold. Ill. J. Math. **25**(4), 555–576 (1981)

[ends3] F. Quinn, Ends of maps. III. Dimensions 4 and 5. J. Differ. Geom. **17**(3), 503–521 (1982)

[fintstern90] R. Fintushel, R.J. Stern, Instanton homology of Seifert fibred homology three spheres. Proc. Lond. Math. Soc. (3) **61**(1), 109–137 (1990)

[fq] M.H. Freedman, F. Quinn, *Topology of 4-Manifolds*. Princeton Mathematical Series, vol. 39 (Princeton University Press, Princeton, 1990)

[freedmanTaylor] M.H. Freedman, L. Taylor, Λ-splitting 4-manifolds. Topology **16**(2), 181–184 (1977)

[gaykirby] D.T. Gay, R. Kirby, *Trisecting 4-manifolds* (2012) arXiv:1205.1565 [math.GT]

[gs] D.E. Galewski, R.J. Stern, Simplicial triangulations of topological manifolds, in *Algebraic and Geometric Topology (Proceedings of Symposia in Pure Mathematics, Stanford University, Stanford, 1976), Part 2*. Proceedings of Symposia in Pure Mathematics, vol. XXXII (American Mathematical Society, Providence, 1978), pp. 7–12

[gs5] D. Galewski, R. Stern, A universal 5-manifold with respect to simplicial triangulations, in *Geometric Topology (Proceedings of the Georgia Topology Conference, Athens, 1977)* (Academic Press, New York, 1979), pp. 345–350

[hist] F. Quinn, A revolution in mathematics? What really happened a century ago and why it matters today. Not. Am. Math. Soc. **59**(1), 31–37 (2012)

[kirby] R.C. Kirby, *The Topology of 4-Manifolds*. Lecture Notes in Mathematics, vol. 1374 (Springer, Berlin, 1989)

[ks] R.C. Kirby, L.C. Siebenmann, *Foundational Essays on Topological Manifolds, Smoothings, and Triangulations* (Princeton University Press, Princeton, 1977). With notes by John Milnor and Michael Atiyah, Annals of Mathematics Studies, No. 88

[kscharlemann79] R.C. Kirby, M.G. Scharlemann, Eight faces of the Poincaré homology 3-sphere, in *Geometric Topology (Proceedings of the Georgia Topology Conference, Athens, 1977)* (Academic Press, New York, 1979), pp. 113–146

[mano03] C. Manolescu. Seiberg-Witten-Floer stable homotopy type of three-manifolds with $b_1 = 0$. Geom. Topol. **7**, 889–932 (electronic) (2003)

[mano13] C. Manolescu, Pin(2)-equivariant Seiberg-Witten Floer homology and the triangulation conjecture. J. Amer. Math. Soc. **29**(1), 147–176 (2016)

[may86] L.G. Lewis Jr., J.P. May, M. Steinberger, J.E. McClure, *Equivariant Stable Homotopy Theory*. Lecture Notes in Mathematics, vol. 1213. (Springer, Berlin, 1986). With contributions by J.E. McClure

[milnor56] J. Milnor, On manifolds homeomorphic to the 7-sphere. Ann. Math. (2) **64**, 399–405 (1956)

[newman26] M.H.A. Newman, The foundations of combinatory analysis situs, i, ii. Proc. Nederl. Akad. Wetensch. **29**, 611–626 (1926) [newman28] M.H.A. Newman, Topological equivalence of complexes. Math. Ann. **99**(1), 399–412 (1928)

[Q1] F. Quinn, An obstruction to the resolution of homology manifolds. Mich. Math. J. **34**(2), 285–291 (1987)

[rev] F. Quinn, A revolution in mathematics? What really happened a century ago and why it matters today. Not. Am. Math. Soc. **59**(1), 31–37 (2012)

[rotvdh] T.O. Rot, R.C.A.M. Vandervorst, Morse-Conley-Floer homology. J. Topol. Anal. **6**(3), 305–338 (2014)

[russell] B. Russell, *Introduction to Mathematical Philosophy*, 2nd edn. (Dover Publications Inc., New York, 1993)

[sullSing] D. Sullivan, Singularities in spaces, in *Proceedings of Liverpool Singularities Symposium, II (1969/1970)*. Lecture Notes in Mathematics, vol. 209 (Springer, Berlin, 1971), pp. 196–206

[tietze] H. Tietze, On the topological invariants of multidimensional manifolds. J. Differ. Equ. **19**, 1–118 (1908)

[tietzeTrans] H. Tietze, On the topological invariants of multidimensional manifolds, translation by John Stillwell. http://www.maths.ed.ac.uk/~aar/haupt/Tietze1908.pdf

[veblenAlex12] O. Veblen, J.W. Alexander II, Manifolds of n dimensions. Ann. Math. (2) **14**(1–4), 163–178 (1912/1913)

[vebWhitehead32] O. Veblen, J.H.C. Whitehead, *The Foundations of Differential Geometry* (Cambridge University Press, Cambridge, 1932)

[whitehead40] J.H.C. Whitehead, On C^1-complexes. Ann. Math. (2) **41**, 809–824 (1940)

[wilder] R.L. Wilder, Development of a theory for manifolds, in *Summary of Lectures and Seminars, Summer Institute on Set-Theoretic Topology Madison 1955* (American Math Society, Berlin, 1957), pp. 7–14

[wilson80] F.W. Wilson Jr., Special structure for C^r monotone Lyapunov functions. J. Differ. Equ. **36**(1), 1–11 (1980)

[wilsonyorke73] F.W. Wilson Jr., J.A. Yorke, Lyapunov functions and isolating blocks. J. Differ. Equ. **13**, 106–123 (1973)

Elliptic Calabi–Yau Threefolds over a del Pezzo Surface

Simon Rose and Noriko Yui

Abstract We consider certain elliptic threefolds over the projective plane (more generally over certain rational surfaces) with a section in Weierstrass normal form. In particular, over a del Pezzo surface of degree 8, these elliptic threefolds are Calabi–Yau threefolds. We will discuss especially the generating functions of Gromov–Witten and Gopakumar–Vafa invariants.

Keywords del Pezzo surface • Calabi–Yau threefold • Modular forms

2000 *Mathematics Subject Classification. Primary*: 14N10, 11F11, 14H52

1 Introduction

During a visit to Max-Planck-Institut für Mathematik Bonn in the spring of 2004, Professor Hirzebruch showed the second author a specific construction of Calabi–Yau threefolds, which are elliptic threefolds over a del Pezzo surface of degree 8 in Weierstrass normal form, that is a family of elliptic curves over a del Pezzo surface of degree 8 (a rational surface) [H], although the construction was known previously in [KMV]. The purpose of this short note is to discuss the generating functions of Gromov–Witten and Gopakumar–Vafa invariants.

This paper was completed while both authors were in residence at the Fields Institute for the thematic program on Calabi–Yau varieties; Arithmetic, Geometry, and Physics which ran from July–December 2013. We thank the hospitality of the Fields Institute. Furthermore, we would like to thank the referee for their comments on the article.

S. Rose
Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100, Copenhagen Ø, Denmark
e-mail: simon.cfr@gmail.com

N. Yui (✉)
Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada K7L 3N6
e-mail: yui@mast.queensu.ca

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_12

2 A del Pezzo Surface of Degree 8

First we will give a definition of a del Pezzo surface. A good reference is Manin [M].

Definition 2.1. A *del Pezzo* surface S is a smooth projective geometrically irreducible surface whose anti-canonical bundle is ample, i.e., $-K_S$ is ample.

The *degree* of S is a positive integer defined by

$$\deg S := K_S \cdot K_S.$$

That is, the degree of S is the self-intersection of its canonical class.

Remark 2.2.

(1) Every del Pezzo surface is geometrically rational. Therefore, it is birationally equivalent to the projective plane, $\mathbb{P}^2$.
(2) Let S be a del Pezzo surface. Then $1 \leq \deg S \leq 9$.
(3) If $\deg S > 2$, then its anti-canonical bundle $-K_S$ is very ample.

Here is a classification results of del Pezzo surfaces according to their degrees.

Theorem 2.3. *Let S be a del Pezzo surface.*

(a) If $\deg S = 4$*, S is birationally equivalent to a complete intersection of two quadrics in* $\mathbb{P}^4$.
(b) If $\deg S = 3$*, S is birationally equivalent to a cubic surface in* $\mathbb{P}^3$.
(c) If $\deg S = 2$*, S is birationally equivalent to a hypersurface of degree* 4 *in the weighted projective* 2*-space* $\mathbb{P}(2,1,1,1)$.
(d) If $\deg S = 1$*, S is birationally equivalent to a hypersurface of degree* 6 *in the weighted projective* 2*-space* $\mathbb{P}(3,2,1,1)$.
(e) Any smooth surface as in (a),(b),(c) or (d) is del Pezzo surface of the expected degree.
(f) Let $P_1, P_2, \cdots, P_r$ *with* $r \leq 8$ *be generic points in* $\mathbb{P}^2$*. Let* $S := Br_{P_1,\cdots,P_r}(\mathbb{P}^2)$ *be the blow-up of* $\mathbb{P}^2$ *at* P_i, $1 \leq i \leq r$*. Then S is a del Pezzo surface of degree* $9 - r$.

To obtain a del Pezzo surface of degree 8, we blow-up $\mathbb{P}^2$ in one point.

Corollary 2.4. *Pick a point* $P \in \mathbb{P}^2$*, and a line* $H \subset \mathbb{P}^2$ *not passing through P. Then*

$$-K_{\mathbb{P}^2} = 3H, \qquad \textit{and} \qquad K_{\mathbb{P}^2} \cdot K_{\mathbb{P}^2} = 9H^2 = 9.$$

Let $S := Bl_P(\mathbb{P}^2)$ *be the blow-up of* $\mathbb{P}^2$ *at P. Furthermore, let E denote the exceptional curve replacing P; then* $E \cdot E = -1$*. Let* $\xi : S \to \mathbb{P}^2$ *be the blow-up map. Then*

$$K_S = \xi^*(K_{\mathbb{P}^2}) + E$$

and

$$\begin{aligned} K_S \cdot K_S &= \xi^*(K_{\mathbb{P}^2}) \cdot \xi^*(K_{\mathbb{P}^2}) + 2\xi^*(K_{\mathbb{P}^2}) \cdot E + E \cdot E \\ &= K_{\mathbb{P}^2} \cdot K_{\mathbb{P}^2} + E \cdot E \\ &= 9 - 1 = 8. \end{aligned}$$

Then S is a del Pezzo surface of degree 8.

Remark 2.5. Let S be a del Pezzo surface of degree d. Then

(1) Every irreducible curve on S is exceptional.
(2) If S has no exceptional curves, then either $d = 9$ and S is isomorphic to $\mathbb{P}^2$, or $d = 8$ and S is isomorphic to $\mathbb{P}^1 \times \mathbb{P}^1$.
(3) If S is not isomorphic to $\mathbb{P}^1 \times \mathbb{P}^1$, then the Picard group $Pic(S)$ is isomorphic to $\mathbb{Z}^{10-d}$. In particular, if $d = 8$, $Pic(S) \simeq \mathbb{Z}^2$ and is spanned by H and E.

3 The Construction of Elliptic Threefolds over S

Let $\pi : X \to S$ be an elliptic fibration, and let L be a line bundle on S with $L \cdot L = 8$. Take

$$g_2 \in H^0(S, L^4), \text{ and } g_3 \in H^0(S, L^6),$$

i.e.,

$$g_2 = 4L \quad \text{and} \quad g_3 = 6L$$

and let

$$X : y^2 z = 4x^3 - g_2 x z^2 - g_3 z^3.$$

Then the canonical bundle K_X is given by

$$K_X = \pi^*(K_{X/S} + K_S) \quad \text{with } K_{X/S} \simeq L^{-1}.$$

We want X to be a Calabi–Yau threefold. The Calabi–Yau condition imposes that

$$K_X \simeq \mathcal{O}_X \Longleftrightarrow K_S = L^{-1} \Longleftrightarrow -K_S = L.$$

Now

$$K_S = -3H + E \Longleftrightarrow L = 3H - E$$

so that

$$4L = 4(3H - E) \qquad \text{and} \qquad 6L = 6(3H - E).$$

Let $[z_0 : z_1 : z_2]$ be the projective coordinate for $\mathbb{P}^2$. Then $g_2 = g_2(z_0, z_1, z_2) \in 4L$ and is of degree 12. While $g_3 = g_3(z_0, z_1, z_2) \in 6L$ and is of degree 18. Put $\Delta = 4g_2^3 - 27g_3^2$. Then $\Delta = \Delta(z_0, z_1, z_2) \in 12L$ and is of degree 36.

4 Calculation of the Euler Characteristic and the Hodge Numbers

Let X be an elliptic threefold constructed above. Then by the construction, the geometric genus of X is $p_g(X) = 1$ and $h^{1,0}(X) = h^{2,0}(X) = 0$. So X is a Calabi–Yau threefold. Now we calculate the Euler characteristic $e(X)$ of X.

Lemma 4.1. *Let Y be a complex surface (possibly with singularities). Then the Euler characteristic $e(Y)(D)$ for any divisor D is given by*

$$e(Y)(D) = K_Y \cdot D - D \cdot D + \textit{Contribution from singularities}.$$

In particular, if Y is smooth,

$$e(Y)(D) = 2 - 2g(Y)$$

which is independent of a choice of a divisor D.

Proposition 4.2. *Let $X : y^2z = 4x^3 - g_2xz^2 - g_3z^3$ be a Calabi–Yau threefold over a del Pezzo surface S, and let $\Delta = 4g_2^3 - 27g_3^2$. Then the Euler characteristic $e(X)$ of X is given by the formula*

$$e(X) = e(\Delta) + \#cusps$$

where the Euler characteristic $e(\Delta)$ of $\{\Delta = 0\}$ is given by

$$e(\Delta) = -degK_\Delta = 2 - 2g(\Delta) + 2\#cusps$$

where $g(\Delta)$ denotes the genus of $\{\Delta = 0\}$.

Moreover, we can compute that

$$\#cusps = 192 \quad \textit{and} \quad g(\Delta) = 595.$$

Finally, we obtain

$$e(X) = -480.$$

Proof. First recall that $L = -K_S$ and that $L \cdot L = 8$. Then we have

$$\begin{aligned} K_\Delta &= (K_S + \Delta) \cdot \Delta = (-L + 12L) \cdot 12L \\ &= 11L \cdot 12L = (11 \cdot 12)(L \cdot L) = 11 \cdot 12 \cdot 8 = 1056. \end{aligned}$$

The number of cusps is given by

$$4L \cdot 6L = 24(L \cdot L) = 24 \cdot 8 = 192.$$

Then

$$e(\Delta) = -1056 + 2 \cdot 192 = -1056 + 384 = -672.$$

Now we need to calculate the Euler characteristic of resolutions of singularities. If $\{\Delta = 0\}$ is smooth, its resolution is an elliptic curve E, and the Euler characteristic $e(E) = 0$. If $\{\Delta = 0\}$ is a node, the Euler characteristic of its resolution is 1, and if $\{\Delta = 0\}$ is a cusp, the Euler characteristic of its resolution is 2.

Then we have

$$\begin{aligned} e(X) &= \left(\begin{array}{c} e(E) \times e(\mathbb{P}^2 \setminus \Delta) \\ +e(\Delta \setminus \{\text{cusps}\}) \times e(\text{resolution of a node}) \\ +\#\text{cusps} \times e(\text{resolution of a cusp}) \end{array} \right) \\ &= e(\Delta) - \#\text{cusps} + 2\#\text{cusps} = e(\Delta) + \#\text{cusps}. \end{aligned}$$

Finally we obtain

$$e(X) = -672 + 192 = -480.$$

□

The Hodge numbers $h^{1,1}(X)$ and $h^{2,1}(X)$ have been calculated by Hulek and Kloosterman [HK] (Sect. 11). This is done by calculating the Mordell–Weil rank of the elliptic curve $\pi : X \to S$, which turns out to be 0.

Lemma 4.3.

$$h^{1,1}(X) = 3, \quad \textit{and} \quad h^{2,1}(X) = 243.$$

The topological Euler characteristic is $e(X) = -480$.

Thus the Hodge diamond is given by

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 0 & & 0 & & \\ & 0 & & 3 & & 0 & \\ 1 & & 243 & & 243 & & 1 \\ & 0 & & 3 & & 0 & \\ & & 0 & & 0 & & \\ & & & 1 & & & \end{array} \qquad \begin{array}{l} B_0(X) = 1 \\ B_1(X) = 0 \\ B_2(X) = 3 \\ B_3(X) = 488 \\ B_4(X) = 3 \\ B_5(X) = 0 \\ B_6(X) = 1 \end{array}$$

Recall X is defined by a Weierstrass equation over the del Pezzo surface S of degree 8 which is birational to $\mathbb{P}^2$,

$$y^2 z = 4x^3 - g_2 x z^2 - g_3 z^3 \quad \text{where } g_2, g_3 \in \mathbb{C}(S)$$

the j-invariant of X is defined by

$$j = 1728\frac{g_2^3}{\Delta} \quad \text{where} \quad \Delta = 4g_2^3 - 27g_3^2.$$

As for these elliptic threefolds, we have

Lemma 4.4. *The j-invariant is a moduli for X.*

We are also interested in the modularity question for the Galois representation associated with X. However, the Betti number $B_3(X) = 488$ is too large to make this practical. Thus, we are interested in constructing a topological mirror Calabi–Yau threefold $\check{X}$.

For a topological mirror partner $\check{X}$ of our elliptic Calabi–Yau threefold X, the Hodge numbers are

$$h^{1,1}(\check{X}) = 243,\ h^{2,1}(\check{X}) = 3$$

and the Euler characteristic is

$$e(\check{X}) = 480.$$

The Betti numbers are

$$B_2(\check{X}) = 243,\ B_3(\check{X}) = 8.$$

In this case, the modularity of the Galois representation may at least somewhat be tractable. This leads us to ask: How can we construct such a mirror Calabi–Yau threefold?

The Calabi–Yau threefold X with the Hodge numbers $h^{1,1}(X) = 3$ and $h^{2,1}(X) = 243$ can be realized in terms of a hypersurface in a toric variety; in fact, it may be realized as a degree 24 hypersurface in weighted projective space with weight $(1, 1, 2, 8, 12)$. Then Batyrev's mirror construction yields 1572 admissible weights which yield not only one but 1572 mirror Calabi–Yau threefolds $\check{X}$ with the Hodge numbers $h^{1,1}(\check{X}) = 243$ and $h^{2,1}(\check{X}) = 3$.

5 Gromov–Witten and Gopakumar–Vafa Invariants

We are naturally interested in the Gromov–Witten invariants of the threefold X. These are obtained via integration against the virtual fundamental class of the moduli space of stable maps into X. That is, we define

$$N^X_{g,\beta} = \int_{[M_{g,n}(X,\beta)]^{\text{vir}}} 1.$$

In the best of cases, these invariants are positive integers and count the number of curves in X in the homology class β. In many cases, however, since $M_{g,n}(X,\beta)$ is a stack, the invariants are only rational numbers.

Naturally, we organize these invariants into a generating function as follows. Let $F_g^X(q)$ and $F^X(q,\lambda)$ be defined as

$$F_g^X(q) = \sum_{\beta\in H_2(X)} N_{g,\beta}^X q^\beta$$

$$F^X(q,\lambda) = \sum_{g=0}^{\infty} \lambda^{2g-2} F_g^X(q).$$

We can now define the *Gopakumar–Vafa/BPS* invariants via the equality

$$F^X(q,\lambda) = \sum_{g=0}^{\infty} \sum_{\beta\in H_2(X)} n_{g,\beta}^X \sum_{m=1}^{\infty} \frac{1}{k}\left(2\sin\left(\frac{k\lambda}{2}\right)\right)^{2g-2} q^{k\beta}.$$

For example, the $g=0$ portion of this reads

$$N_{0,\beta}^X = \sum_{\substack{\eta\in H_2(X)\\ k\eta=\beta}} \frac{1}{k^3} n_{0,\eta}^X.$$

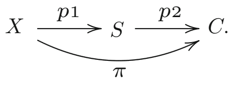
These invariants $n_{g,\beta}^X$ are defined recursively in terms of the Gromov–Witten invariants $N_{g,\beta}^X$, and a priori these are only rational numbers. It is a conjecture (see [GV, K]) that they are integers for all X, g, β. We work with them because in the case of the Calabi–Yau threefold X, the formulæ for them turn out to be much simpler; the Gromov–Witten invariants can then be reconstructed from them.

In the case that a class β is primitive, the invariants $N_{0,\beta}^X$ and $n_{0,\beta}^X$ coincide.

6 The Geometry of X

In order to compute these invariants, we need a bit more of a description of the geometry of the threefold X. We begin with the following fact. The del Pezzo surface S is in fact isomorphic to the Hirzebruch surface $\mathbb{F}_1 = \mathbb{P}(\mathcal{O}\oplus\mathcal{O}(1))$. This is a $\mathbb{P}^1$ bundle over the base $\mathbb{P}^1$, with a (-1)-curve as a section. Let C', F' denote the homology classes in S of the section and fibre, respectively.

Consider now the following composition



The generic fibre of this is an elliptically fibred K3 surface with 24 I_1 fibres.

Let now X_F denote one such generic fibre, and let X_C denote the restriction of X to the section C. This latter surface is a rational elliptic surface with 12 I_1 fibres (which physicists call a $\frac{1}{2}$K3). Similarly, let C'', E'' denote the class of the section and fibre in X_C, respectively.

We want to have a description of the Picard group and a basis of $H_2(X, \mathbb{Z})$. So consider first the line bundles

$$L_1 = \mathcal{O}(S) \qquad L_2 = \mathcal{O}(X_C) \qquad L_3 = \mathcal{O}(X_F)$$

and let $\iota_1, \iota_2, \iota_3$ denote the respective inclusions of S, X_C, X_F. We now define the homology classes

$$\begin{aligned} C &= (\iota_1)_*(C') = (\iota_2)_*(C'') \\ E &= (\iota_2)_*(E'') \\ F &= (\iota_1)_*(F'). \end{aligned}$$

Lemma 6.1. *The line bundles L_1, L_2, L_3 form a basis of the Picard group of X, and the classes C, E, F form a basis of $H_2(X, \mathbb{Z})$ (which are all effective).*

Proof. We can compute the intersection pairing of these bundles with these curves, which we find to be

	C	F	E
L_1	-1	-2	1
L_2	-1	1	0
L_3	1	0	0

which clearly has determinant -1. It follows (since $h^{1,1} = 3$) that the lattices that these generate must be the whole lattice. □

We will further need the triple intersections $\Gamma_{ijk} = \int_X L_i \smile L_j \smile L_k$, which are computed as follows.

Lemma 6.2. *The triple intersections are given by the following.*

$$\begin{aligned} &\Gamma_{111} = 8 && \Gamma_{112} = -1 && \Gamma_{113} = -2 \\ &\Gamma_{122} = -1 && \Gamma_{123} = 1 && \Gamma_{133} = 0 \\ &\Gamma_{222} = 0 && \Gamma_{223} = 0 && \Gamma_{233} = 0 \\ & && \Gamma_{333} = 0 && \end{aligned}$$

Proof. These are computed simply by restricting the line bundles to the smooth representatives S, X_C, X_F, where the intersections are easy to compute. □

6.1 The Rational Elliptic Surface

The rational elliptic surface X_C is realizable as the blowup of $\mathbb{P}^2$ at 9 points in general position; as such, its intersection form is $\Gamma_{1,9}$, where $\Gamma_{a,b}$ is the lattice with diagonal intersection form given by

$$\mathrm{diag}(\underbrace{1,\ldots,1}_{a},\underbrace{-1,\ldots,-1}_{b}).$$

Let $H, C_0, \ldots, C_8$ denote the classes of the line and exceptional curves, respectively. Then $3H - \sum_{i=0}^{8} C_i$ and C_0 span a sublattice which is isomorphic to $\Gamma_{1,1}$ (of discriminant 1), and hence we have a splitting

$$\Gamma_{1,9} \cong \Gamma_{1,1} \oplus E_8$$

where E_8 is the unique even, unimodular, negative-definite lattice corresponding to the Dynkin diagram E_8. We should remark that the classes C_0 and $3H - \sum_{i=0}^{8} C_i$ are the same as the base and fibre classes C'', E'' of the rational elliptic surface discussed earlier.

Remark 6.3. The canonical divisor on the surface X_C is given by

$$K_{X_C} = -3H + \sum_{i=0}^{8} C_i = -E''.$$

This allows us now to compute the relationship between the groups $H_2(X_C, \mathbb{Z})$ and $H_2(X, \mathbb{Z})$, which we will need later.

Lemma 6.4. *The map $H_2(X_C, \mathbb{Z}) \to H_2(X, \mathbb{Z})$ is given by*

$$H_2(X_C, \mathbb{Z}) \cong \Gamma_{1,1} \oplus E_8 \xrightarrow{proj} \Gamma_{1,1} \subseteq H_2(X, \mathbb{Z})$$

where $\Gamma_{1,1}$ includes via the identification $C := (\iota_2)_* C'', E := (\iota_2)_*(E'')$ described earlier.

Proof. We first claim that $(\iota_2)_* C_i = C + E$ for $1 \leq i \leq 8$, and that $(\iota_2)_* H = 3(C + E)$. This can be seen simply by using the push–pull formula and noting that

$$\iota_2^* L_1 = C_0 \qquad \iota_2^* L_2 = -E'' \qquad \iota_2^* L_3 = E''.$$

Now, an element $aH + \sum_{i=0}^{8} b_i C_i$ is in the orthogonal complement of the lattice spanned by $E'' = 3H - \sum_{i=0}^{8} C_i, C'' = C_0$ if and only if

(1) $b_0 = 0$
(2) $3a + \sum_{i=0}^{8} b_i = 0.$

Thus, we have that

$$(\iota_2)_* \left(aH + \sum_{i=0}^{8} b_i C_i \right) = 3a(C+E) + \sum_{i=0}^{8} b_i (C+E)$$
$$= \left(3a + \sum_{i=0}^{8} b_i \right)(C+E) = 0$$

as claimed. □

Finally, we need one fact about effectivity of classes in $H_2(X_C, \mathbb{Z})$.

Lemma 6.5. *Let $\beta = C'' + nE'' + \lambda \in H_2(X_C, \mathbb{Z})$, where $\lambda \in E_8$. Then β is effective if and only if $\lambda \cdot \lambda \geq -2n$.*

Proof. This is a straightforward application of Riemann–Roch. For a divisor D on X_C, this reads as

$$\chi(D) = 1 + \frac{1}{2} D \cdot (D - K_{X_C}).$$

In particular, for $D = C'' + nE'' + \lambda$, we find that

$$h^0(D) + h^2(K_{X_C} - D) = 1 + n + \frac{1}{2}\lambda \cdot \lambda.$$

Thus since $K_{X_C} - D$ will never be effective, it follows that as long as $n + \frac{1}{2}\lambda \cdot \lambda \geq 0$, that D will have a section, and hence be effective. □

6.2 The K3 Fibration

To compute the Gopakumar–Vafa invariants of X in the fibre-wise classes (i.e., those which project down to 0 under the map $\pi : X \to C$), we use the machinery of [KMPS, MP], which we will review here. Moreover, the ideas in this section closely follow the ideas of [KMPS]. For more detail, that article is strongly recommended.

Definition 6.6. Let Λ be a rank r lattice. A family of Λ-polarized K3 surfaces over a base curve Σ is a scheme Z over Σ together with a collection of line bundles $L_1, \ldots, L_r$ such that, for each $b \in \Sigma$, the fibre $(X_b, L_1|_{X_b}, \ldots, L_r|_{X_b})$ is a Λ-polarized K3 surface.

Such a family $Z \xrightarrow{\pi} \Sigma$ yields a map ι_π to the moduli space $\mathcal{M}_\Lambda$ of Λ-polarized K3 surfaces. Intersecting the image of the curve with certain divisors in $\mathcal{M}_\Lambda$ (see again, [KMPS, MP]) will produce the *Noether–Lefschetz numbers*. These are given as follows.

The Noether–Lefschetz divisors consist of those Λ-polarized K3 surfaces which jump in Picard rank; these are determined by

(1) an integer h, such that the square of the new class β is $2h-2$
(2) r integers $d_1,\ldots,d_r$ which are given by $d_i = \int_\beta L_i$.

We denote such a divisor by $D_{h;d_1,\ldots,d_r}$, and we then define

$$NL^{\pi}_{(h;d_1,\ldots,d_r)} = \int_{\iota_\pi \Sigma} D_{h;d_1,\ldots,d_r}.$$

It should be remarked that, by the Hodge index theorem, this will be only be non-zero if the discriminant

$$\Delta(h;d_1,\ldots,d_r) = (-1)^r \det \begin{pmatrix} \boxed{\Lambda} & \begin{matrix} d_1 \\ \vdots \\ d_r \end{matrix} \\ d_1 \cdots d_r & 2h-2 \end{pmatrix}$$

is non-negative.

Let $r_{0,h}$ denote the reduced Gopakumar–Vafa invariants of a K3 surface in a class β such that $\beta\cdot\beta = 2h-2$. From [KMPS], these only depend on the square of β (and not its divisibility, as one might expect), and they satisfy the *Yau–Zaslow formula* (see [BL, G, KMPS, YZ])

$$\sum_{h=0}^{\infty} r_{0,h} q^{h-1} = \frac{1}{\Delta(q)} = \frac{1}{\eta(q)^{24}} = q^{-1} + 24 + 324q + 3200q^2 + \cdots.$$

It should be remarked that the power of 24 that shows up in this formula is due to the presence of the 24 nodal fibres in our elliptically fibred K3 surfaces.

Let $n^Z_{(d_1,\ldots,d_r)}$ denote the Gopakumar–Vafa invariants of the threefold Z defined by

$$n^Z_{(d_1,\ldots,d_r)} = \sum_{\substack{\beta\in H_2(Z) \\ \int_\beta L_i = d_i}} n^Z_\beta.$$

We have the following relation between these invariants.

Theorem 6.7 ([MP, Theorem 1*]). *The invariants $n^Z_{(d_1,\ldots,d_r)}$, $r_{0,h}$, and $NL^\pi_{h;d_1,\ldots,d_r}$ satisfy the following relationship.*

$$n^Z_{(d_1,\ldots,d_r)} = \sum_{h=0}^{\infty} r_{0,h} NL^\pi_{h;d_1,\ldots,d_r}$$

Consider now the restriction of L_1, L_2 to X_F. We can compute their intersections via Lemma 6.2 to find that we have

$$\begin{pmatrix} L_1 \cdot L_1 & L_1 \cdot L_2 \\ L_2 \cdot L_1 & L_2 \cdot L_2 \end{pmatrix} = \underbrace{\begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix}}_{\Lambda}$$

where in this case the rank r of Λ is 2.

What we would like to have is that the triple (X, L_1, L_2) is a family of Λ-polarized K3 surfaces. However, due to the presence of singular fibres (due to the singularities of $\Delta = 4g_2^3 - 27g_3^2$) this is not the case. However, we can "resolve" this threefold (see [KMPS, MP]) to obtain a threefold $\tilde{X} \xrightarrow{\tilde{\pi}} C$ which is such a family. We can then relate the invariants of the two families as follows, allowing us to compute the Gopakumar–Vafa invariants of X as desired.

Lemma 6.8. *The invariants of $X, \tilde{X}$ satisfy*

$$n^{\tilde{X}}_{(d_1,d_2)} = 2n^X_{(d_1,d_2)}.$$

Our final ingredient is to note that the Noether-Lefschetz numbers are coefficients of a modular form of weight $\frac{22-r}{2} = 10$; that is, they are the coefficients of some multiple $E_4(z)E_6(z) = E_{10}(z) = 1 - 264q - 135432q^2 - \cdots$. Thus we need to only compute a single such coefficient to determine all of the Noether-Lefschetz numbers.

Definition 6.9. Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$. Then we will use the notation

$$[n]f(z) = a_n$$

to denote the coefficient of z^n in $f(z)$.

Using this notation, we have the following lemma.

Lemma 6.10. *We have that*

$$NL^{\tilde{X}}_{0;0,0} = 1056$$

and so consequently we have that

$$NL^{\tilde{\pi}}_{h;d_1,d_2} = -4\left[\frac{\Delta(h;d_1,d_2)}{2}\right]E_{10}(z).$$

Proof. The proof of this is identical to the proofs of Lemma 2 and Proposition 2 of [KMPS], and thus we omit it. □

7 Computations of the Generating Functions

We are now ready to compute the generating functions for the Gopakumar–Vafa invariants. The generating functions we are interested are those of the following form.

Choose $\beta = mC' + rF' \in H_2(S)$ (which we will identify from now on for simplicity's sake with its image in $H_2(X, \mathbb{Z})$). Define

$$F_\beta(q) = \sum_{n=0}^{\infty} n^X_{\beta+nE} q^{n-m-\frac{1}{2}r}.$$

Remark 7.1. We choose this shift in the exponent of q to match the results in [KMW]. This ensures that the generating functions that we obtain below will be modular, but we don't have a better interpretation of this shifted power.

We then have the following theorem.

Theorem 7.2. *We have the following expressions for generating functions $F_\beta(q)$:*

$$F_F(q) = -2\frac{E_{10}(q)}{\Delta(q)} = -2q^{-1} + 480 + 282888q + 17058560q^2 +, \cdots$$
$$F_C(q) = \frac{E_4(q)}{\sqrt{\Delta(q)}} = q^{-\frac{1}{2}} + 252q^{\frac{1}{2}} + 5130q^{\frac{3}{2}} + 54760q^{\frac{5}{2}} + \cdots$$

Each of these is a meromorphic modular form of weight −2, and moreover each of the generating functions $F_{mF}(q^m)$ is also (meromorphic) modular of the same weight, but for the group $\Gamma_1(m^2)$.

The first two of these generating functions are conjectured (with physical justification) in the papers [KMW, KMV], along with a few others (in particular, [AS], where these arise as their initial conditions to compute the higher genus topological string amplitudes). We have not found any prior description of the third, although it is an easy generalization of the first.

We will split the proof of this theorem up into several parts.

Theorem 7.3. *We have the equality*

$$F_F(q) = -2\frac{E_{10}(q)}{\Delta(q)}.$$

Proof. We will compute first the function $F_F(q)$. Since we have that the class $F+nE$ is determined uniquely by its integration against L_1, L_2, we have that

$$n^X_{F+nE} = n^X_{(n-2,1)}.$$

Combining Lemmata 6.10, 6.8, and Theorem 6.7, our generating function $F_F(q)$ is given by

$$\begin{aligned} F_F(q) &= \sum_{n=0}^{\infty} n^{X}_{(n-2,1)} q^{n-1} \\ &= \sum_{n=0}^{\infty} \tfrac{1}{2} n^{\tilde{X}}_{(n-2,1)} q^{n-1} \\ &= \frac{1}{2} \sum_{n=0}^{\infty} \sum_{h=0}^{\infty} r_{0,h} NL^{\tilde{\pi}}_{h;n-2,1} q^{n-1}. \end{aligned}$$

We can compute the discriminant $\Delta(h; n-2, 1) = 2n - 2h$ which must be non-negative, so the summation is really over those $n, h \geq 0$ with $n \geq h$. Thus we can write this as

$$\begin{aligned} F_F(q) &= \frac{1}{2} \sum_{h=0}^{\infty} r_{0,h} \sum_{n=h}^{\infty} NL^{\tilde{\pi}}_{h;n-2,1} q^{n-1} \\ &= \frac{1}{2} \sum_{h=0}^{\infty} r_{0,h} q^{h-1} \sum_{n=h}^{\infty} (-4) \left[\frac{2n-2h}{2} \right] E_{10}(z) q^{n-h} \\ &= -2 \sum_{h=0}^{\infty} r_{0,h} q^{h-1} \sum_{n=h}^{\infty} [n-h] E_{10}(z) q^{n-h} \\ &= -2 \sum_{h=0}^{\infty} r_{0,h} q^{h-1} E_{10}(q) \\ &= -2 \frac{E_{10}(q)}{\Delta(q)}. \end{aligned}$$

□

To prove the next formula, we need the following computation of the Gromov–Witten invariants (for primitive classes) of a rational elliptic surface.

Theorem 7.4 ([BL], Theorem 6.2). *The generating function for the Gopakumar–Vafa invariants of the rational elliptic surface X_C in the classes $C'' + nE''$ is given by*

$$\sum_{n=0}^{\infty} n^{X_C}_{C''+nE''} q^{n-\frac{1}{2}} = \frac{1}{\sqrt{\Delta(q)}}.$$

We now prove the following.

Theorem 7.5. *We have the equality*

$$F_C(q) = \frac{E_4(q)}{\sqrt{\Delta(q)}}.$$

Proof. Recall from Lemma 6.4 that the map $(\iota_2)_* : H_2(X_C, \mathbb{Z}) \cong \Gamma_{1,1} \oplus E_8 \to H_2(X, \mathbb{Z})$ is essentially the projection onto the $\Gamma_{1,1}$ factor.

Now, we obtain curves in X in the class $C + nE$ by considering curves in X_C in some effective class which pushes forward to this class; from Lemma 6.4, these will be curves of the form $\beta = C'' + nE'' + \lambda$ where $\lambda \in E_8$. From Lemma 6.5, we know that the effective ones are those with $n \geq -\frac{1}{2}\lambda \cdot \lambda$.

We can now compute that

$$
\begin{aligned}
F_C(q) &= \sum_{n=0}^{\infty} n^X_{C+nE} q^{n-\frac{1}{2}} \\
&= \sum_{n=0} \sum_{\substack{\beta \in H_2(X_C) \\ (\iota_2)_*\beta = C+nE}} n^{X_C}_{\beta} q^{n-\frac{1}{2}} \\
&= \sum_{n=0}^{\infty} \sum_{\substack{\lambda \in E_8 \\ -\lambda\cdot\lambda \leq 2n}} n^{X_C}_{C''+nE''+\lambda} q^{n-\frac{1}{2}}.
\end{aligned}
$$

Now, morally similar to the case of K3 surfaces, from primitive curve classes the invariants $n^{X_C}_{\beta}$ only depend on the square of β. In particular, any such curve can be transformed into one of the form $\beta = C'' + nE''$ by a series of Cremona transformations and permutations of the exceptional classes (see [GP]). It follows then that

$$
n^{X_C}_{C''+nE''+\lambda} = n^{X_C}_{C''+\left(n+\frac{1}{2}\lambda\cdot\lambda\right)E''}
$$

and so the generating function becomes

$$
\begin{aligned}
F_C(q) &= \sum_{n=0}^{\infty} \sum_{\substack{\lambda \in E_8 \\ -\lambda\cdot\lambda \leq 2n}} n^{X_C}_{C_0+(n+\frac{1}{2}\lambda\cdot\lambda)E} q^{n-\frac{1}{2}} \\
&= \sum_{\lambda \in E_8} q^{-\frac{1}{2}\lambda\cdot\lambda} \sum_{n=-\frac{1}{2}\lambda\cdot\lambda}^{\infty} n^{X_C}_{C_0+(n+\frac{1}{2}\lambda\cdot\lambda)E} q^{n+\frac{1}{2}\lambda\cdot\lambda-\frac{1}{2}} \\
&= \sum_{\lambda \in E_8} q^{-\frac{1}{2}\lambda\cdot\lambda} \sum_{n=0}^{\infty} n^{X_C}_{C_0+nE} q^{n-\frac{1}{2}} \\
&= \Theta_{E_8}(q)\left(\frac{1}{\sqrt{\Delta(q)}}\right) = \frac{E_4(q)}{\sqrt{\Delta(q)}}
\end{aligned}
$$

as claimed (with the last equality being due to the well-known fact that $\Theta_{E_8}(q) = E_4(q)$). □

To prove the last statement in the theorem, we need a little extra notation.

Definition 7.6. Let $f(z) = \sum a_n z^n$ be a power series, and let m, k be integers with $0 \leq k < m$. We define then

$$f_{m,k}(z) = \sum_{\substack{n \equiv k \\ (\text{mod } m)}} a_n z^n = \sum_{n=0}^{\infty} a_{mn+k} z^{mn+k}.$$

We should note that in the case that $f(z)$ is a modular form of weight r for $SL_2(\mathbb{Z})$, then each of the functions $f_{m,k}(z)$ is also modular of the same weight for the subgroup $\Gamma_1(m^2)$.

Furthermore, we can expand this definition for values of k outside of the given range by replacing k with a suitable integer congruent to $k \pmod m$ within that range. For example, $f_{m,-1}(z) = f_{m,m-1}(z)$.

We can now state more precisely our final theorem.

Theorem 7.7. *Let $m > 1$. Then the generating function $F_{mF}(q)$ is given by*

$$F_{mF}(q) = \sum_{n=0}^{\infty} n_{mF+nE} q^{n-m} = -2 \sum_{\ell=0}^{m-1} \left(\frac{1}{\Delta(u)} \right)_{m,\ell-1} \left(E_{10}(u)\right)_{m,1-\ell}$$

where $q = u^m$.

Proof. This proof follows very similarly to the proof of Theorem 7.3. We similarly begin with noting that $n^X_{mF+nE} = n^X_{(n-2m,m)}$, which allows us to write

$$\begin{aligned} F_{mF}(q) &= \sum_{n=0}^{\infty} n^X_{(n-2m,m)} q^{n-m} \\ &= \sum_{n=0}^{\infty} \tfrac{1}{2} n^{\tilde{X}}_{(n-2m,m)} q^{n-m} \\ &= \frac{1}{2} \sum_{n=0}^{\infty} \sum_{h=0}^{\infty} r_{0,h} N L^{\tilde{\pi}}_{h;n-2m,m} q^{n-m}. \end{aligned}$$

In this case, the discriminant $\Delta(h; n-2m, m) = 2 - 2h + 2nm - 2m^2$ which as usual must be non-negative, leaving us summing over all pairs (h, n) such that $n \geq m + \frac{h-1}{m}$. Thus we obtain

$$F_{mF}(q) = -2 \sum_{h=0}^{\infty} r_{0,h} \sum_{n \geq m + \frac{h-1}{m}} [1 - h + nm - m^2] E_{10}(z) q^{n-m}.$$

To simplify this further, we split the summation over h into a sum over congruence classes mod m. If we let $q = u^m$, then this yields the following.

$$
\begin{aligned}
F_{mF}(q) &= -2\sum_{\ell=0}^{m-1}\sum_{h=0}^{\infty} r_{0,mh+\ell} \sum_{n\geq m+h+\frac{\ell-1}{m}} [1-\ell+m(n-h-m)]E_{10}(z)u^{nm-m^2} \\
&= -2\sum_{\ell=0}^{m-1}\sum_{h=0}^{\infty} r_{0,mh+\ell}u^{mh+\ell-1} \sum_{n\geq m+h+\frac{\ell-1}{m}} [1-\ell+m(n-h-m)]E_{10}(z)u^{m(n-h-m)-\ell+1} \\
&= -2\sum_{\ell=0}^{m-1}\sum_{h=0}^{\infty} r_{0,mh+\ell}u^{mh+\ell-1}(E_{10})_{m,1-\ell}(u) \\
&= -2\sum_{\ell=0}^{m-1}\left(\frac{1}{\Delta(u)}\right)_{m,\ell-1}\big(E_{10}(u)\big)_{m,1-\ell}
\end{aligned}
$$

which ends the proof. □

The above results show that we end up with meromorphic modular forms when we consider generating functions for Gopakumar–Vafa invariants for curve classes of the form $mF + nE$ and $C + nE$. From the conjectured results in [KMW], it seems that we should end up with similar results for curve classes of the form $rC + nE$; a natural approach to study these would be to use the recursion of [GP], which we will look to do at a future date.

Acknowledgements S. Rose was supported by a Jerry Marsden Postdoctoral Fellowship for the Fields major thematic program on Calabi–Yau Varieties: Arithmetic, Geometry and Physics from July to December 2013. N. Yui was supported in part by the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant.

References

[AS] M. Alim, E. Scheidegger, Topological strings on elliptic fibrations. Commun. Number Theory Phys. **8** (2012). doi:10.4310/CNTP.2014.v8.n4.a4

[BL] J. Bryan, N.C. Leung, The enumerative geometry of $K3$ surfaces and modular forms. J. Am. Math. Soc. **13**(2), 371–410 (electronic) (2000). MR1750955 (2001i:14071)

[GV] R. Gopakumar, C. Vafa, M-theory and topological strings I and II. (1998). arXiv:hep-th19809187; hep-th19812127

[GP] L. Göttsche, R. Pandharipande, The quantum cohomology of blow-ups of $\mathbf{P}^2$ and enumerative geometry. J. Differ. Geom. **48**(1), 61–90 (1998). 1622601 (99d:14057)

[G] L. Göttsche, A conjectural generating function for numbers of curves on surfaces. Commun. Math. Phys. **196**(3), 523–533 (1998). 1645204 (2000f:14085)

[H] F. Hirzebruch, Classical algebraic geometry and Calabi-Yau manifolds, in *Calabi Conference at University of Wisconsin* (2003)

[HK] K. Hulek, R. Kloosterman, Calculating the Mordell-Weil rank of elliptic threefolds and the cohomology of singular hypersurfaces. Ann. Inst. Fourier (Grenoble) **61**(3), 1133–1179 (2011). 2918726

[K] S. Katz, Gromov-Witten, Gopakumar-Vafa, and Donaldson-Thomas invariants of Calabi-Yau threefolds, in *Snowbird Lectures on String Geometry*. Contemporary Mathematics, vol. 401 (American Mathematical Society, Providence, RI, 2006), pp. 43–52. 2222528 (2007b:14121)

[KMPS] A. Klemm, D. Maulik, R. Pandharipande, E. Scheidegger, Noether-Lefschetz theory and the Yau-Zaslow conjecture. J. Am. Math. Soc. **23**(4), 1013–1040 (2010). 2669707 (2011j:14121)

[KMV] A. Klemm, P. Mayr, C. Vafa, BPS states of exceptional non-critical strings. Nucl. Phys. B Proc. Suppl. **58**, 177–194 (1997). Advanced Quantum Field Theory (La Londe les Maures, 1996). 1486340 (99a:81145)

[KMW] A. Klemm, J. Manschot, T. Wotschke, Quantum geometry of elliptic Calabi-Yau manifolds. Commun. Number Theory Phys. **6**(4), 849–917 (2012). 3068410

[M] Y. Manin, *Cubic Forms: Algebra, Geometry, Arithmetic*. North-Holland Mathematical Library (Elsevier Science, Amsterdam, 1986)

[MP] D. Maulik, R. Pandharipande, Gromov-Witten theory and Noether-Lefschetz theory, in *A Celebration of Algebraic Geometry*, ed. by B. Hassett, J. McKernan, J. Starr, R. Vakil. Clay Mathematics Proceedings, vol. 18 (American Mathematical Society, Providence, RI, 2013), pp. 469–508

[YZ] S.-T. Yau, E. Zaslow, BPS states, string duality, and nodal curves on $K3$. Nucl. Phys. B **471**(3), 503–512 (1996). MR1398633 (97e:14066)

Remarks on Cohomological Hall Algebras and Their Representations

Yan Soibelman

Dedicated to the memory of F. Hirzebruch

1 Introduction

The aim of this paper is to discuss a class of representations of Cohomological Hall algebras related to the notion of framed stable object of a category. The paper is an extended version of the talk the author gave at the workshop on Donaldson–Thomas invariants at the University Paris-7 in June 2013 and at the conference "Algebra, Geometry, Physics" dedicated to Maxim Kontsevich (June 2014, IHES). Because of the origin of the paper it contains more speculations than proofs.

1.1 Cohomological Hall Algebras and Their Representations: Motivations

The notion of Cohomological Hall algebra (COHA for short) for quivers with potential was introduced in [KoSo5].[1] Since a quiver with potential defines a 3-dimensional Calabi–Yau category ($3CY$ category for short), it was expected that COHA could be defined for "good" subcategories of 3-dimensional Calabi–Yau categories endowed with additional data (most notably, orientation data introduced

[1] In fact we considered in the loc. cit. more general case of formally smooth algebras with potential.

Y. Soibelman (✉)
Department of Mathematics, KSU, Manhattan, KS 66506, USA
e-mail: soibel@math.ksu.edu

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_13

in [KoSo1]), or, more generally, for an ind-constructible locally regular $3CY$ category in the sense of [KoSo1]. At the time of writing this general goal has not been achieved, although there have been some progress in special cases.[2]

The definition of COHA given in [KoSo5] is similar to the definition of conventional (constructible) Hall algebra (see, e.g., [Sch1]) or its motivic version (see [KoSo1]). Instead of spaces of constructible functions on the stack of objects of an abelian category we considered in [KoSo5] cohomology groups. Certain correspondences define "structure constants" of the multiplication. There are many versions of COHA depending on a choice of "cohomology theory" (see [KoSo5]). It is expected that there is an upgrade of COHA to a dg-algebra in the (triangulated version of the) category of exponential motives (see [KoSo5]). Then all other versions will appear as "realizations" of this dg-algebra.

By analogy with conventional Hall algebra of a quiver, which gives a quantization of the "positive" part of the corresponding Lie algebra, one may want to define "full" (or "double") COHA, for which the one defined in [KoSo5] will be just the "positive part." At this time we do not know the comultiplication which makes COHA into a bialgebra. Hence we cannot follow Drinfeld double construction (which works in the case of constructible Hall algebras).

On the other hand, we can try to define full COHA by means of representation theory, similarly to the classical approach of Nakajima to the infinite Heisenberg algebras (see [Nak2]). In this way one hopes to reconstruct full COHA from its representation theory.

One of the motivations for COHA comes from supersymmetric Quantum Field Theory and String Theory, where spaces of BPS states can be often identified with the cohomology groups of various moduli spaces. From this perspective COHA can be thought of as a mathematical implementation of the idea of BPS algebra (see [HaMo1, HaMo2]). Representation theory of BPS algebras has not been developed by physicists, although such a theory should have interesting applications. Furthermore, various dualities in physics can lead to natural mathematical questions about COHA, otherwise unmotivated. For example, our approach to COHA is based on 3-dimensional Calabi–Yau categories. The latter appear in the geometric engineering story on the string-theoretic side. Taking seriously the idea that COHA (or its double) is the BPS algebra, one can ask about the corresponding structures on the gauge-theoretic (i.e., "instanton") side of the geometric engineering. There is some interesting mathematical work related to this question (see, e.g., [Nak1, SchV, So1, Sz2]).

In any case the author believes that the representation theory of COHA should be developed further in order to approach some of the above-mentioned problems.

[2]For example, I heard from Dominic Joyce about the work in progress of Oren Ben-Bassat in which COHA would be defined for the derived moduli stack of perfect complexes on a compact Calabi–Yau three-fold.

1.2 COHA and Sheaf of Vanishing Cycles

Below we will give a very schematic description of COHA assuming that the above-mentioned "good" category is abelian.

Let $\mathbf{k}$ be a perfect field. Suppose $\mathcal{C}$ is a $\mathbf{k}$-linear triangulated A_∞-category, which is ind-constructible and locally regular in the sense of Sect. 3 of [KoSo1].[3] It is explained in Sect. 3.2 of loc. cit. that one can associate with $\mathcal{C}$ the ind-constructible stack $\mathcal{M}_\mathcal{C}$ of objects of $\mathcal{C}$. Local regularity implies that $\mathcal{M}_\mathcal{C}$ is locally presented as an ind-Artin stack over $\mathbf{k}$. Let $\mathcal{A} \subset \mathcal{C}$ be an abelian subcategory. Then we have an ind-constructible substack $\mathcal{M}_\mathcal{A} \subset \mathcal{M}_\mathcal{C}$ of objects of $\mathcal{A}$, which is locally ind-Artin.

Hypothetical definition of COHA depends on an ind-constructible sheaf Φ on $\mathcal{M}_\mathcal{C}$ (more precisely, a perverse sheaf). In the case of $3CY$ categories one takes $\Phi = \phi_W$, which is the sheaf of vanishing cycles of the potential W (we recall the definition of the potential in Sect. 2.1). For the sheaf of vanishing cycles to be well-defined on $\mathcal{M}_\mathcal{C}$, the $3CY$-category $\mathcal{C}$ has to be endowed with an *orientation data*. The latter is an ind-constructible super line bundle $\mathcal{L}$ over $\mathcal{M}_\mathcal{C}$, such that for the fiber over a point of $\mathcal{M}_\mathcal{C}$ corresponding to an object $E \in \mathrm{Ob}(\mathcal{C})$ one has $\mathcal{L}_E^{\otimes 2} = s\det(\mathrm{Ext}^\bullet(E,E))$. Furthermore, it is required that $\mathcal{L}_E$ behaves naturally on exact triangles (see [KoSo1], Sect. 5 for the details). It follows from local regularity that $\mathcal{L}$ is (locally) a line bundle over an ind-Artin stack.

Let $i : \mathcal{M}_\mathcal{A} \subset \mathcal{M}_\mathcal{C}$ be the natural embedding. Then the pull-back $i^*(\Phi)$ is an ind-constructible sheaf on $\mathcal{M}_\mathcal{A}$. Let $\mathcal{Z} \subset \mathcal{M}_\mathcal{A} \times \mathcal{M}_\mathcal{A}$ be the "Hecke correspondence," which is the stack consisting of pairs of objects (E,F) such that $E \subset F$. There are projections $p_n : \mathcal{Z} \to \mathcal{M}_\mathcal{A}, n = 1,2,3$ such that $p_1(E,F) = E, p_2(E,F) = F, p_3(E,F) = E/F$. One of the properties that we require from "good" abelian category $\mathcal{A}$ is that the projection p_2 is a locally proper morphism of ind-Artin stacks.

As a vector space COHA of $\mathcal{M}_\mathcal{A}$ is defined as $\mathcal{H} := \mathcal{H}_\mathcal{A} = \mathbf{H}^\bullet(\mathcal{M}_\mathcal{A}, i^*(\Phi))$. For that one chooses an appropriate cohomology theory $\mathbf{H}^\bullet$ of Artin stacks with coefficients in constructible sheaves. The product $m : \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{A} \to \mathcal{H}_\mathcal{A}$ is defined by the formula $p_{2*} \circ (p_1^* \otimes p_3^*)(i^*(\Phi) \otimes i^*(\Phi))$.

For the sheaf of vanishing cycles ϕ_W associativity of the product depends on the Thom–Sebastiani theorem for the chosen cohomology theory.[4]

Remark 1.2.1. As a part of the data we fix an abstract version of Chern character (called the class map in [KoSo1]). It is a homomorphism of abelian groups $\mathrm{cl} : K_0(\mathcal{C}) \to \Gamma$, where $\Gamma \simeq \mathbf{Z}^n$ is a free abelian group, such that connected components of $\mathcal{M}_\mathcal{C}$ are parametrized by Γ, while classes $\mathrm{cl}(E), E \in \mathrm{Ob}(\mathcal{A})$ form an additive submonoid $\Gamma_+ \subset \Gamma$. Then COHA of $\mathcal{A}$ will be Γ_+-graded algebra.

[3] Calabi–Yau structure which we will discuss later leads to the requirement $\mathrm{char}(\mathbf{k}) = 0$. For simplicity we will often assume that $\mathbf{k} = \mathbf{C}$.

[4] As explained in Sect. 7 of [KoSo5], it is more convenient to work with compactly supported cohomology and then apply the duality functor.

In the case of smooth algebras with potential considered in [KoSo5] the stack $\mathcal{M}_{\mathcal{A}}$ is a countable union of smooth quotient stacks, and the foundational questions are resolved positively. For some ideas about the general case one can look at [DyKap]. Natural class of examples for which COHA should be defined comes from the theory of graded symplectic manifolds (see [PaToVaVe] about foundational material and [BuJoMe] about some progress in the definition of ϕ_W).

1.3 Stable Framed Objects and Modules over COHA

It is natural to ask whether one can realize representations of COHA in the cohomology groups of some natural schemes (or stacks), which might also depend on a choice of stability condition on $\mathcal{A}$. Let us explain how it can be achieved.[5]

First we define the moduli space $\mathcal{M}_\gamma^{\mathrm{fr,st}}, \gamma \in \Gamma_+$ of " stable framed objects of class γ" (in applications those can be framed stable sheaves, framed stable representations of quivers, framed special Lagrangian submanifolds, etc.). This notion depends on a choice of stability condition on $\mathcal{A}$. It is expected (see [KoSo7]) that the sheaf ϕ_W "descends" to each moduli space $\mathcal{M}_\gamma^{\mathrm{fr,st}}$.

For a pair of classes $\gamma_1, \gamma_2 \in \Gamma_+$ let us consider the Hecke correspondence $\mathcal{Z}_{\gamma_1,\gamma_2}$ of pairs $(E_{\gamma_1+\gamma_2}, E_{\gamma_2})$ (the subscripts denote the Chern classes) of framed stable objects such that E_{γ_2} is a quotient of $E_{\gamma_1+\gamma_2}$. Let us denote the cohomology theory we used in the definition of COHA by $\mathbf{H}^\bullet$. It descends to each $\mathcal{M}_\gamma^{\mathrm{fr,st}}$. Furthermore, similarly to the definition of COHA we have three projections of $\mathcal{Z}_{\gamma_1,\gamma_2}$:

(a) to $\mathcal{M}_{\gamma_2}^{\mathrm{fr,st}}$;
(b) to the moduli space $\mathcal{M}_{\gamma_1}$ of all (not framed) objects with fixed γ_1;
(c) to $\mathcal{M}_{\gamma_1+\gamma_2}^{\mathrm{fr,st}}$.

Using the pull-back and pushforward construction as in the previous subsection, we obtain a structure of $\mathcal{H}_{\mathcal{A}} = \oplus_\gamma \mathbf{H}^\bullet(\mathcal{M}_\gamma, \phi_W)$-module over COHA of $\mathcal{A}$ on the space $\oplus_\gamma \mathbf{H}^\bullet(\mathcal{M}_\gamma^{\mathrm{fr,st}}, \phi_W)$.

We are going to discuss stable framed objects in Sect. 3. We will show there that the moduli stacks of stable framed objects are in fact schemes. Hence graded components of our representations of COHA are *finite-dimensional* vector spaces. As a version of the above considerations we can drop the stability assumption and consider stacks of framed (but not necessarily stable framed) objects. Then we still obtain representations of COHA. But this time the graded components of the representation spaces will be in general infinite-dimensional.

[5]We warn the reader that even in the case of quivers our moduli spaces are **not** Nakajima quiver varieties.

1.4 Modules over COHA Motivated by Physics

As we have already mentioned, COHA can be thought of as a mathematical incarnation of the algebra of (closed) BPS states envisioned in [HaMo1, HaMo2]. According to [KoSo5] (refined) BPS-invariants can be computed as virtual Poincaré polynomials of graded components of COHA. One can ask about the meaning of the algebra structure on COHA.

Motivated by the ideas of S. Gukov (see, e.g., [GuSto]) we would like to think about representations of COHA described in the previous subsection as of representations of the algebra of closed BPS states on the vector space of open BPS states. We are going to speculate about applications of this point of view in the last section of the paper. We plan to discuss the relationship between COHA and BPS-algebras more systematically in separate projects jointly with E. Diaconescu, S.Gukov, N. Saulina.

Here we just mention three interesting classes of representations of COHA which have geometric origin and should have interesting applications to gauge theory and knot invariants:

(a) Representation of COHA of the resolved conifold $X = \mathrm{tot}(\mathcal{O}_{\mathbf{P}^1}(-1) \oplus \mathcal{O}_{\mathbf{P}^1}(-1))$ realized in the cohomology of moduli spaces of C-framed stable sheaves in the sense of [DiHuSo]. Those modules should have applications in the theory of algebraic knots and Hilbert schemes of curves (see [ORS, DiHuSo, Mau1]).
(b) Modules over COHAs of the Fukaya categories of non-compact Calabi–Yau three-folds associated with spectral curves of Hitchin integrable systems. Those should serve as BPS algebras of some gauge theories from the class $\mathcal{S}$ (see, e.g., [GaMoNe-2, Ga1] about the latter).
(c) This class of examples was already mentioned above. Based on the ideas of geometric engineering we hope for a class of representations of COHA related to the moduli spaces of (framed, possibly ramified) instantons on $\mathbf{P}^2$. One can hope to understand the relationship between the algebras of Hecke operators proposed in [Nak1] and those proposed in [So1]. Currently we can see COHA on the "Calabi–Yau side" of geometric engineering and the (seemingly unrelated) conventional ("motivic") Hall algebra on the "instanton side" (cf., also [SchV, Sz2]).

The relationship between various classes of gauge theories might give non-trivial results about corresponding COHAs and their representations (including the relations between (a), (b), (c)).

1.5 Contents of the Paper

Section 2 is a reminder on COHA in the framework of quivers with potential. Section 3 is devoted to stable framed objects in triangulated and abelian categories. In Sect. 4 we discuss representations of COHA realized in the cohomology of the moduli spaces of stable framed representations. We also discuss an approach to the definition of "full COHA" based on the representation theory in the cohomology of moduli of stable framed objects. In Sect. 5 we speculate about representations of COHA motivated by knot theory and physics.

2 Cohomological Hall Algebra: Reminder

This section is a reminder of some basic facts about the notion of Cohomological Hall algebra. Most of the material is borrowed from [KoSo5], and we refer the reader to loc. cit. for more details and proofs.

2.1 COHA and 3CY Categories

Suppose we are given an ind-constructible locally regular $3CY$ category $\mathcal{C}$ over the field $\mathbf{k}, \mathrm{char}(\mathbf{k}) = 0$ (see [KoSo1]). As explained in Sect. 3.2 of loc. cit., one can associate with such a category the stack of objects, which is a countable disjoint union of schemes over $\mathbf{k}$ of finite type acted by affine algebraic groups. For simplicity of the exposition we take the ground field $\mathbf{k} = \mathbf{C}$.

Some examples of such categories are listed in the Introduction of [KoSo1]. They include various categories of D-branes popular in string theory (e.g., the Fukaya category of a compact or local Calabi–Yau three-fold, the category of perfect sheaves on such a three-fold, the category of finite-dimensional representations of a quiver with potential, etc.).

In order to define COHA one has to choose *orientation data* (see [KoSo1], Sect. 5) on $\mathcal{C}$ as well as a "good" t-structure with the ind-Artin heart. Let us denote it by $\mathcal{A}$. The existence of mutation-invariant orientation data is known for a class of $3CY$ categories associated with a quiver without potential (see [Dav1]). There are partial existence results for the derived category of coherent sheaves on a compact Calabi–Yau three-fold (see, e.g., [Hu]). Probably Dominic Joyce with collaborators will construct soon an orientation data on the stack of objects of the category of perfect complexes on a Calabi–Yau three-fold. But the general case is still open. In the present paper we will assume the existence of the orientation data as a part of the "foundational" package. Also, we do not discuss in detail the meaning of the notion of "good" t-structure. As we mentioned in the Introduction, the latter includes properness of the morphisms which appear in the definition of the product on COHA.

We assume as part of the data the "class map" $\mathrm{cl} : K_0(\mathcal{C}) \to \Gamma$ (see [KoSo1]), where $\Gamma \simeq \mathbf{Z}^n$ is a free abelian group endowed with integer skew-symmetric form $\langle \bullet, \bullet \rangle$ (Poisson lattice). We also assume that the class map respects the Euler form $\chi(E,F) = \sum_i (-1)^i \dim \mathrm{Ext}^i(E,F)$ on $K_0(\mathcal{C})$ and the form $\langle \bullet, \bullet \rangle$ on Γ. The lattice Γ

plays a role of topological K-theory of the category $\mathcal{C}$. Finally, we assume that we have fixed an additive submonoid $\Gamma_+ \subset \Gamma$ generated by $\mathrm{cl}(E), E \in \mathrm{Ob}(\mathcal{A})$.

When the above choices are made, one can define COHA of $\mathcal{A}$ as an associative algebra graded by Γ_+. Graded components are given by the cohomology of the moduli stacks of objects with the given class $\gamma \in \Gamma$ with the coefficients in the sheaf of vanishing cycles of the potential of $\mathcal{C}$ restricted to $\mathcal{A}$.

For completeness we recall here the notion of potential of a $3CY$ category. Using the A_∞-structure on $\mathcal{C}$ as well as the Calabi–Yau pairing $(\bullet, \bullet)$ (see [KoSo1]) one defines the potential of an object E as a formal series:

$$W_E(a) = \sum_{n \geq 1} \frac{(m_n(a, \ldots, a), a)}{n+1},$$

where m_n are higher composition maps, and the element a belongs to $\mathrm{Hom}^1(E, E)$ which is the subspace in the graded space $\mathrm{Hom}(E, E)$ consisting of elements of degree 1. By our assumptions the potential W_E is a locally regular function with respect to E. Hence we have a partially formal function W defined by the family of series W_E.

Remark 2.1.1. If $\mathcal{C}$ is "minimal on the diagonal" (see [KoSo1]), we can replace $\mathrm{Hom}(E, E)$ by its cohomology with respect to the differential m_1. In this case we may assume that $a \in \mathrm{Ext}^1(E, E)$, which can be thought of as the "tangent space to the moduli stack of formal deformations of E". Hence one can think of the potential as a function on the moduli stack of objects which is locally regular along the stack of objects (this follows from the "locally regular" assumption) and formal in the transversal direction.

Then COHA is a Γ-graded vector space

$$\mathcal{H} := \oplus_{\gamma \in \Gamma} \mathcal{H}_\gamma \ ,$$

where $\mathcal{H}_\gamma = H^\bullet_{\mathsf{G}_\gamma}(S_\gamma, W_\gamma)$, and S_γ is the stack of objects E such that $\mathrm{cl}(E) = \gamma$. Recall that we use an appropriate stack version of the cohomology theory $H^\bullet(X, f)$ of a scheme X endowed with a regular function f. There are several choices for such theory. They are discussed in [KoSo5], where the above approach made rigorous in the case of $3CY$ categories arising from quivers (more generally, formally smooth algebras) with potential. A version of the cohomology theory which is suitable in the framework of categories is called "critical cohomology" in loc. cit. It is defined by means of the compactly supported cohomology of X with coefficients in the sheaf of vanishing cycles of f. Sometimes (e.g., for quivers with potential) the function $f := W$ is regular. In such a case one can use de Rham cohomology defined via the twisted de Rham differential $d + dW \wedge (\bullet)$ or Betti cohomology which is generated by "integration cycles" for the exponential differential forms of the type $exp(W)\nu$. More generally, one can define "motivic" version of COHA. In that case COHA $\mathcal{H}$ is an object of the tensor category of exponential mixed Hodge structures, and the concrete choice of the cohomology theory corresponds to a tensor functor to graded vector spaces ("realization"). It is explained in [KoSo5] that in all realizations $\mathcal{H}$ carries an associative algebra structure with "structure constants" defined by means of the cohomology of certain "Hecke correspondences" with coefficients in the sheaves of vanishing cycles of the potential $W = (W_\gamma)_{\gamma\in\Gamma}$.

Let us illustrate the above considerations in the case of a quiver Q with potential W, which is the main example in [KoSo5]. We set $\mathbf{k} = \mathbf{C}$. If I is the set of vertices of Q, then $\Gamma = \mathbf{Z}^I$, $\Gamma_+ = \mathbf{Z}^I_{\geq 0}$. For any $\gamma = (\gamma^i)_{i\in I} \in \Gamma_+$ we consider γ-dimensional representations of Q in coordinate vector spaces $(\mathbf{C}^{\gamma^i})_{i\in I}$. It is an affine scheme M_γ naturally acted by the affine algebraic group $\mathsf{G}_\gamma = \prod_{i\in I} \mathrm{GL}(\gamma^i, \mathbf{C})$. Then the corresponding stack of objects is a countable union (over all dimension vectors $\gamma \in \Gamma_+$) of algebraic varieties $\mathrm{Crit}(W_\gamma)$ of the critical points of the functions $W_\gamma = \mathrm{Tr}(W) : \mathsf{M}_\gamma \to \mathbf{C}$. Then COHA is the direct sum $\oplus_{\gamma\in\Gamma_+} H^\bullet_{\mathsf{G}_\gamma}(\mathsf{M}_\gamma, W_\gamma)$ with the product defined in the loc. cit. In the next three subsections we are going to recall more explicit descriptions of the product in some examples.

2.2 COHA for Quivers Without Potential

COHA is non-trivial even if $W = 0$. In the latter case

$$\mathcal{H}_\gamma := H^\bullet_{\mathsf{G}_\gamma}(\mathsf{M}_\gamma).$$

Since M_γ is equivariantly contractible, and the quotient of each $\mathrm{GL}(\gamma^i, \mathbf{C})$ by the normalizer of the maximal torus is homotopy equivalent to the point, one can use the toric localization and obtain an explicit formula for the product which expresses COHA as a shuffle algebra. In the formula below we identify equivariant cohomology of a point with respect to the trivial action of the group $GL(\gamma^i, \mathbf{C})$ with the space of symmetric polynomials in the corresponding group of γ^i variables.

Theorem 2.2.1. *The product $f_1 \cdot f_2$ of elements $f_i \in \mathcal{H}_{\gamma_i}$, $i = 1, 2$ is given by the symmetric function $g((x_{i,\alpha})_{i\in I,\alpha\in\{1,\dots,\gamma^i\}})$, where $\gamma := \gamma_1 + \gamma_2$, obtained from the following function in variables $(x'_{i,\alpha})_{i\in I,\alpha\in\{1,\dots,\gamma_1^i\}}$ and $(x''_{i,\alpha})_{i\in I,\alpha\in\{1,\dots,\gamma_2^i\}}$:*

$$f_1((x'_{i,\alpha}))f_2((x''_{i,\alpha}))\ \frac{\prod_{i,j\in I}\prod_{\alpha_1=1}^{\gamma_1^i}\prod_{\alpha_2=1}^{\gamma_2^j}(x''_{j,\alpha_2} - x'_{i,\alpha_1})^{a_{ij}}}{\prod_{i\in I}\prod_{\alpha_1=1}^{\gamma_1^i}\prod_{\alpha_2=1}^{\gamma_2^i}(x''_{i,\alpha_2} - x'_{i,\alpha_1})}\ ,$$

by taking the sum over all shuffles for any given $i \in I$ of the variables $x'_{i,\alpha}, x''_{i,\alpha}$ (the sum is over $\prod_{i\in I}\binom{\gamma^i}{\gamma_1^i}$ shuffles).

Here a_{ij} is the number of arrows in Q from the vertex i to vertex j.

For example, let $Q = Q_d$ be a quiver with just one vertex and $d \geq 0$ loops. Then the product formula specializes to

$$(f_1 \cdot f_2)(x_1, \ldots, x_{n+m}) := \sum_{i_1,\ldots,j_m} f_1(x_{i_1}, \ldots, x_{i_n}) f_2(x_{j_1}, \ldots, x_{j_m}) \left(\prod_{k=1}^{n} \prod_{l=1}^{m} (x_{j_l} - x_{i_k}) \right)^{d-1}$$

for symmetric polynomials, where f_1 has n variables, and f_2 has m variables. The sum is taken over all $\{i_1 < \cdots < i_n, j_1 < \cdots < j_m, \{i_1, \ldots, i_n, j_1, \ldots, j_m\} = \{1, \ldots, n + m\}$. The product $f_1 \cdot f_2$ is a symmetric polynomial in $n + m$ variables. One can show that for even d the algebra is isomorphic to the infinite Grassmann algebra, while for odd d one gets an infinite symmetric algebra.

We introduce a double grading on algebra $\mathcal{H}$, by declaring that a homogeneous symmetric polynomial of degree k in n variables has bigrading $(n, 2k + (1 - d)n^2)$. Equivalently, one can shift the cohomological grading in $H^\bullet(\mathrm{BGL}(n, \mathbf{C}))$ by $[(d - 1)n^2]$. In general, even for quivers without potential each component $\mathcal{H}_\gamma$ has also the grading by cohomological degree. Total $\Gamma \times \mathbf{Z}$-grading can be further refined, since $\mathcal{H}_\gamma$ carries the weight filtration (as an object of the category of exponential mixed Hodge structures, see [KoSo5]). Hence typically COHA has $\Gamma \times \mathbf{Z} \times \mathbf{Z}$-grading (which is not compatible with the product). More precisely, it is shown in [KoSo5] that for $W = 0$ COHA is graded by the Heisenberg group.

Finally, we remark that in the case of Dynkin quivers there are other interesting explicit formulas for the product in COHA (see [Rim]).

2.3 COHA for Quiver A_2

The quiver A_2 has two vertices $\{1, 2\}$ and one arrow $1 \leftarrow 2$. The Cohomological Hall algebra $\mathcal{H}$ of this quiver contains two subalgebras $\mathcal{H}_L$, $\mathcal{H}_R$ corresponding to representations supported at the vertices 1 and 2, respectively. Clearly each subalgebra $\mathcal{H}_L$, $\mathcal{H}_R$ is isomorphic to the Cohomological Hall algebra for the quiver $A_1 = Q_0$. Hence it is an infinite Grassmann algebra. Let us denote the generators by ξ_i, $i = 0, 1, \ldots$ for the vertex 1 and by η_i, $i = 0, 1, \ldots$ for the vertex 2. Each generator ξ_i or η_i corresponds to an additive generator of the group $H^{2i}(\mathrm{BGL}(1, \mathbb{C})) \simeq \mathbb{Z} \cdot x^i$. Then one can check that $\xi_i, \eta_j,\ i, j \geqslant 0$ satisfy the relations

$$\xi_i \xi_j + \xi_j \xi_i = \eta_i \eta_j + \eta_j \eta_i = 0\,, \quad \eta_i\, \xi_j = \xi_{j+1} \eta_i - \xi_j \eta_{i+1}\,.$$

Let us introduce the elements $\nu_i^1 = \xi_0 \eta_i\,,\ i \geqslant 0$ and $\nu_i^2 = \xi_i \eta_0\,,\ i \geqslant 0$. It is easy to see that $\nu_i^1 \nu_j^1 + \nu_j^1 \nu_i^1 = 0$, and similarly the generators ν_i^2 anticommute. Thus we have two infinite Grassmann subalgebras in $\mathcal{H}$ corresponding to these two choices: $\mathcal{H}^{(1)} \simeq \bigwedge(\nu_i^1)_{i \geqslant 0}$ and

$\mathcal{H}^{(2)} \simeq \bigwedge(\nu_i^2)_{i \geqslant 0}$. One can directly check the following result.

Proposition 2.3.1. *The multiplication (from the left to the right) induces isomorphisms of graded vector spaces:*

$$\mathcal{H}_L \otimes \mathcal{H}_R \xrightarrow{\sim} \mathcal{H}, \quad \mathcal{H}_R \otimes \mathcal{H}^{(i)} \otimes \mathcal{H}_L \xrightarrow{\sim} \mathcal{H}\,,\ i = 1, 2\,.$$

2.4 COHA for Jordan Quiver with Polynomial Potential

Let us consider the quiver Q_1 which has one vertex and one loop l (Jordan quiver), and choose as the potential $W = \sum_{i=0}^{N} c_i l^i,\ \ c_N \neq 0$ an arbitrary polynomial of degree $N \in \mathbf{Z}_{\geqslant 0}$ in one variable.

In the case $N = 0$, the question about COHA reduces to the quiver Q_1 without potential. This case was considered before. The algebra $\mathcal{H}$ is the symmetric algebra of infinitely many variables.

In the case $N = 1$ COHA is one-dimensional.

In the case $N = 2$ we may assume without loss of generality that $W = -l^2$. Then COHA $\mathcal{H} = \mathcal{H}^{(Q_1,W)}$ is the exterior algebra with infinitely many generators (infinite Grassmann algebra). This can be shown directly.

In the case when the degree $N \geqslant 3$, one can show that the bigraded algebra $\mathcal{H}$ is isomorphic to the $(N-1)$-st tensor power of the infinite Grassmann algebra of the case $N = 2$.

Basically the above examples are the only cases in which we know COHA explicitly. On the other hand, generating functions for the dimensions of its graded components (we call them *motivic DT-series* in [KoSo1, KoSo5]) are known in many cases.

2.5 Stability Conditions and Motivic DT-Invariants

Definition of COHA depends on the abelian category $\mathcal{A}$ but does not depend on the central charge, which is a homomorphism of groups $Z : \Gamma \to \mathbf{C}$. This raises the question about the role of Bridgeland stability condition in the structure of COHA.

Let us fix a stability condition on a $3CY$-category $\mathcal{C}$. We understand the stability condition in the sense of [KoSo1] (it differs from the conventional Bridgeland's axiomatics by axiomatizing the "class map" $\mathrm{cl} : K_0(\mathcal{C}) \to \Gamma \simeq \mathbf{Z}^n$ and by the so-called Support Property axiom). According to Bridgeland, a choice of stability condition is equivalent to a choice of t-structure and a central charge. Let $\mathcal{A}$ be the heart of the t-structure, and $Z : \Gamma \to \mathbf{C}$ be the central charge. Then for each strict sector $V \subset \mathbf{R}^2$ with the vertex at the origin we can define a full subcategory $\mathcal{A}_V \subset \mathcal{A}$ generated (using extensions) by the zero object and semistable objects with the central charge sitting in V. For example, we can take $V = l$ to be a ray. Taking V

to be the upper-half plane we recover $\mathcal{A}$. In these two cases the categories generated by semistables are abelian. For other sectors V it is not the case in general.

As explained in [KoSo5], for a fixed strict sector V, one can define a Γ-graded *vector space*

$$\mathcal{H}(V) := \oplus_{\gamma\in\Gamma}\mathcal{H}_\gamma(V)\ .$$

This space is an algebra only in the cases when $V = l$ or V is an upper-half plane. For other sectors V one has faces the problem of properness of morphisms of the corresponding stacks.

It was observed in [KoSo5], Sect. 5.2 that the algebras $\mathcal{H}_l := \mathcal{H}(l)$ resemble universal enveloping algebras of some Lie algebras $\mathfrak{g}_l$ which are analogous to the "positive root" Lie algebras $\mathfrak{g}_\alpha, \alpha > 0$ of Kac-Moody algebras. Then similarly to the isomorphism $U(n_+) \simeq \otimes_{\alpha>0}U(\mathfrak{g}_\alpha)$ (which depends on a chosen order on the set of positive roots) one should expect an isomorphism $\mathcal{H}(V) \simeq \otimes_{l\subset V}\mathcal{H}_l$ where the tensor product is taken in the clockwise order over all rays in the sector V. This was demonstrated in [Rim] in the case of Dynkin quivers without potential. In particular, taking V to be the upper-half plane we obtain a factorization of the COHA $\mathcal{H}$ into the tensor product of COHAs for individual rays. COHA for each ray l is typically commutative. It can be computed from the knowledge of space of semistable objects in the fixed t-structure whose central charges belong to l. For a generic central charge we have two possibilities: either l does not contain $Z(\gamma)$ for $\gamma \in \Gamma$, or l contains only multiples $nZ(\gamma_0), n > 0$ for some primitive vector γ_0 (an furthermore, only vectors $n\gamma_0, n \in \mathbf{Z}_{>0}$ are mapped by Z to l). In this case $\mathcal{H}_l$ is indeed commutative and can be computed explicitly in many cases.

The notion of motivic DT-series (i.e., virtual Poincaré series of $\mathcal{H}$) does not depend on the central charge. On the other hand, motivic *DT-invariants* $\Omega^{mot}(\gamma)$ (they correspond in physics to refined BPS invariants) can be defined only after a choice of stability condition (i.e., the central charge in case of quivers). Definition of DT-invariants is based on the theory of factorization systems developed in [KoSo5]. It follows from loc. cit. that the motivic DT-series factorizes as a product of the powers of shifted quantum dilogarithms. Those powers are motivic DT-invariants.

As a side remark we mention that factorization systems appear in different disguises when mathematicians try to make sense of the operator product expansion in physics (we can mention, e.g., the work of Beilinson and Drinfeld on chiral algebras or the work of Costello and others on OPE in QFT). From this point of view it is not quite clear why factorization systems appear in our story.

2.6 Generators of COHA

For different t-structures the corresponding COHAs are not necessarily isomorphic. For example, if we start with a pair (Q, W) consisting of a quiver Q with potential W and make a mutation at a vertex $i_0 \in I$, then COHA for the mutated pair (Q', W')

is different from the one for (Q, W). On the other hand, we can compute motivic DT-series for the mutated quiver with potential. As was explained in [KoSo1] and [KoSo5], if we make a mutation at the vertex $i_0 \in I$, then the motivic DT-series for (Q, W) and (Q', W') are related by the conjugation by the motivic DT-series corresponding to the ray $l_0 = \mathbf{R}_{>0} \cdot Z(\gamma_{i_0})$ (which is essentially the quantum dilogarithm).

2.6.0 How to define COHA for a *triangulated* $3CY$ category $\mathcal{C}$?
We do not know the answer to this question, but we can see some structures which should be incorporated in the definition.

For example, let us consider all COHAs corresponding to all possible mutations. Let M be the orbit of the pair (Q, W) under the action of the group of mutations. Then to any $m \in M$ we can assign COHA $\mathcal{H}_m$. More generally, we can consider rotations $Z \mapsto Ze^{i\theta}$ of the central charge and get the corresponding COHA $\mathcal{H}_{e^{i\theta}}$. This defines a structure of cosheaf of algebras over S^1. Each stalk is the COHA for the corresponding t-structure.

Next question is about the space of generators of COHA. Recall the following conjecture from [KoSo5] which was proved by Efimov (see [Ef]). It is formulated for symmetric quivers. Such quivers arise naturally in relation to 2-dimensional Calabi–Yau categories and Kac-Moody algebras.

Theorem 2.6.2. *Let $\mathcal{H}$ be the COHA (considered as an algebra over $\mathbf{Q}$) for the abelian category of finite-dimensional representations of a symmetric quiver Q. Then $\mathcal{H}$ is a free supercommutative algebra generated by a graded vector space V over $\mathbf{Q}$ of the form $V = V' \otimes \mathbf{Q}[x]$, where x is an even variable of bidegree $(0, 2) \in \mathbf{Z}^I_{\geq 0} \times \mathbf{Z}$, and for any given γ the space $V'_{\gamma,k} \neq 0$ is non-zero (and finite-dimensional) only for* ***finitely many*** *$k \in \mathbf{Z}$.*

In general we expect (see [KoSo5] for the precise question) that $\mathcal{H}$ is isomorphic to the universal enveloping algebra of a graded Lie algebra $V := V' \otimes \mathbf{C}[x]$ which satisfies the conditions of the Theorem 2.6.2. Mutations act on V, hence we obtain a collection of vector spaces V_m (one for each t-structure m). From the point of view of chamber structure of the space of stability conditions, we can say that with every chamber we associate its own COHA. Change of the chamber corresponds to the wall-crossing, which at the level of COHA is a conjugation (with a shift of grading). More structural results generalizing on COHA, including generalizations of the above Theorem 2.6.2 and their applications (e.g., to Kac conjecture) can be found in recent papers by B. Davison (see [Dav2, Dav3]).

3 Framed and Stable Framed Objects

In this section we present a definition of stable framed objects following [KoSo7] as well as a related construction of modules over COHA of the same authors (unpublished).

3.1 Stable Framed Objects in Triangulated Categories

We recall the definition of stable framed object from [KoSo7] in the case of triangulated categories. Then we discuss some versions in the case of abelian categories.

Let $\mathcal{C}$ be a triangulated A_∞-category over the ground field $\mathbf{k}$, which we assume to be an algebraically closed of characteristic zero. We fix a stability condition $\tau \in Stab(\mathcal{C})$. Let $\Phi : \mathcal{C} \to D^b(\mathrm{Vect}_{\mathbf{k}})$ be an exact functor to the triangulated category of bounded complexes of $\mathbf{k}$-vector spaces.

For a fixed ray l in the upper-half plane with the vertex at the origin, we denote by $\mathcal{C}_l := \mathcal{C}_l^{ss}$ the abelian category of τ-semistable objects having the central charge in l. We will impose the following assumption: Φ *maps* $\mathcal{C}_l$ *to the complexes concentrated in non-negative degrees*.

Definition 3.1.1. Framed object (or Φ-framed object, if we want to stress dependence on the framing functor) is a pair (E,f) where $E \in \mathrm{Ob}(\mathcal{C}_l)$ and $f \in H^0(\Phi(E))$.

Let (E_1,f_1) and (E_2,f_2) be two framed objects. We define a morphism $\phi : (E_1,f_1) \to (E_2,f_2)$ as a morphism $E_1 \to E_2$ such that the induced map $H^0(\Phi(E_1)) \to H^0(\Phi(E_2))$ maps f_1 to f_2. Framed objects naturally form a category, and hence there is a notion of isomorphic framed objects.

Definition 3.1.2. We call the framed object (E,f) stable is there is no exact triangle $E' \to E \to E''$ in $\mathcal{C}$ with E' non-isomorphic to E such that both $E', E'' \in \mathrm{Ob}(\mathcal{C}_l)$ and such that there is $f' \in H^0(\Phi(E'))$ which is mapped to $f \in H^0(\Phi(E))$.

Then one deduces the following result (see [KoSo7]), proof of which we reproduce here for completeness.

Proposition 3.1.3. *If (E,f) is a stable framed object, then $Aut(E,f) = \{1\}$.*

Proof. Let $h \in Aut(E)$ satisfies the property that its image $\Phi(h)$ preserves f. We may assume that $h \in \mathrm{Hom}^0(E,E)$. We would like to prove that $h = id$. Assume the contrary. Let $h_1 := h - id \neq 0$. Then $\Phi(h_1)(f) = 0$. Since the category $\mathcal{C}_l$ is abelian, the morphism $h_1 \neq 0$ gives rise to a short exact sequence in $\mathcal{C}_l$:

$$0 \to \mathrm{Ker}(h_1) \to E \to \mathrm{Im}(h_1) \to 0,$$

where $\mathrm{Im}(h_1) \neq 0$. Hence there exists an exact triangle $E' \to E \to E''$ in $\mathcal{C}$ with $E' = \mathrm{Ker}(h_1)$ non-isomorphic to E and $E'' = \mathrm{Im}(h_1)$. Let us consider an exact sequence in $\mathcal{C}_l$ given by

$$0 \to \mathrm{Ker}(h_1) \to E \to E \to \mathrm{Coker}(h_1) \to 0,$$

where the morphism $E \to E$ is h_1. Using exactness of the functor Φ one can derive a long exact sequence of vector spaces

$$H^0(\mathrm{Ker}(\Phi(h_1))) \to H^0(\Phi(E)) \to H^0(\Phi(E))$$
$$\to H^0(\mathrm{Ker}(\Phi(h_1))) \to H^1(\mathrm{Ker}(\Phi(h_1))) \to \dots .$$

In any case, we remark that since Φ maps $\mathcal{C}_l$ to complexes with non-negative cohomology, we conclude that if $E' \to E \to E''$ is an exact triangle, then in the induced exact sequence

$$H^{-1}(\Phi(E'')) \to H^0(\Phi(E')) \to H^0(\Phi(E)) \to \dots$$

the first terms are trivial. Hence the functor $H^0\Phi$ maps monomorphisms in $\mathcal{C}_l$ to monomorphisms in the category $\mathrm{Vect}_{\mathbf{k}}$ of $\mathbf{k}$-vector spaces.

Let us decompose h_1 into a composition of the morphism $\psi : E \to \mathrm{Im}(h_1)$ and the natural embedding $j : \mathrm{Im}(h_1) \to E$. Applying Φ, and using $\Phi(h_1)(f) = 0$ and the above remark we conclude that $\Phi(\psi)(f) = 0$.

Finally, applying Φ to the short exact sequence

$$0 \to \mathrm{Ker}(h_1) \to E \to \mathrm{Im}(h_1) \to 0,$$

we obtain a short exact sequence in $\mathrm{Vect}_{\mathbf{k}}$:

$$H^0(\mathrm{Ker}(\Phi(h_1))) \to H^0(\Phi(E)) \to H^0(\Phi(\mathrm{Im}(h_1))),$$

where the last arrow is $\Phi(\psi)$. Since $\Phi(\psi)(f) = 0$ we conclude that there exists $f_1 \in H^0(\mathrm{Ker}(\Phi(h_1)))$ which is mapped into f. This contradicts to the assumption that the pair (E, f) is framed stable. The Proposition is proved. ■

The above Proposition makes plausible the following result:

Corollary 3.1.4. *The moduli stack of stable framed objects is in fact a scheme.*

In many examples it is a smooth projective scheme (cf., [Re1]).

3.2 Stable Framed Objects and Torsion Pairs

The above definitions can be repeated almost word by word, if we replace an ind-Artin (or locally regular) triangulated category $\mathcal{C}$ by an ind-Artin abelian category $\mathcal{A}$. Then we have a definition of the framed and stable framed objects in the framework of abelian categories. Let us discuss its relation to the classical notion of torsion pair (see, e.g., [H] for a short introduction).

Recall that a torsion pair for the abelian category $\mathcal{A}$ is given by a pairs of two full subcategories $\mathcal{T}, \mathcal{F} \subset \mathcal{A}$ such that $\mathrm{Hom}(T, F) = 0$ for any pair $T \in \mathrm{Ob}(\mathcal{T}), F \in \mathrm{Ob}(\mathcal{F})$ and such that any object $E \in \mathrm{Ob}(\mathcal{F})$ admits (a unique) decomposition

$$0 \to T \to E \to F \to 0$$

with the same meaning of F and T. Here T is called the *torsion* part of E and F is called the *torsion-free* part of E. The origin of the terminology is clear from the theory of abelian groups or theory of coherent sheaves on curves.

Let us assume as before that our abelian category $\mathcal{A}$ is $\mathbf{k}$-linear. Suppose we are given a stability condition on $\mathcal{A}$ with the central charge Z. Fix $\theta \in (0,\pi)$. Then the pair of full subcategories $\mathcal{T}_\theta = \{T \in \mathrm{Ob}(\mathcal{A}|\mathrm{Arg}(Z(T)) > \theta\}$, $\mathcal{F}_\theta = \{F \in \mathrm{Ob}(\mathcal{A}|\mathrm{Arg}(Z(F)) \leq \theta\}$ define (under some additional conditions on Harder–Narasimhan subquotients of objects in each of the two categories, see Remark 3.2.2 below)[6] a torsion pair for $\mathcal{A}$ (one can exchange strict and non-strict inequality signs). Let us fix a non-zero object $P \in \mathrm{Ob}(\mathcal{A})$. It defines a functor $\mathcal{F}_\theta \to \mathrm{Vect}_{\mathbf{k}}$ given by $\Phi(E) = \mathrm{Hom}(P,E)$. Framed objects are pairs $(E, f : P \to E)$. Then we can give the following version of the notion of stable framed object: (E,f) *is stable framed if either f is epimorphism or* $\mathrm{Coker}(f)$ *is a non-zero object of* $\mathcal{T}_\theta$.

Then the above Proposition 3.1.3 still holds, and the proof is much simpler.

Proposition 3.2.1. *The automorphism group of a stable framed object is trivial.*

Proof. Let $h : E \to E$ be an automorphism such that $h \circ f = f$. Then $(h - id)$ vanishes on the image of f. If f is an epimorphism, we conclude that $h = id$. Otherwise, assume $h \neq id$. Then $(h - id)$ defines a non-trivial morphism $\mathrm{Coker}(f) \to E$ which contradicts to the assumption on $\mathrm{Coker}(f)$ and the definition of torsion pair. Hence $h = id$. ∎

From this Proposition we again conclude that stable framed objects form a scheme, not a stack.

Remark 3.2.2. Notice that in the proof of Proposition 3.2.1 we did not really use a fixed slope θ, we rather worked with an individual object E. Hence we can give the following version of the notion of stable framed object for the framing functor defined by means of an object P: stable framed object is a pair (E,f) such that E is a non-zero object of category $\mathcal{A}$, $f : P \to E$ is a morphism which is either an epimorphism or a morphism with non-zero cokernel satisfying the condition that $\mathrm{Arg}(\mathrm{Coker}(f)) > \mathrm{Arg}(E)$, and the same inequality holds for all subquotients of these objects (we denote $\mathrm{Arg}(Z(E))$ by $\mathrm{Arg}(E)$ to simplify the notation). Yet another possibility is to require that all Harder–Narasimhan factors of E belong to $\mathcal{T}_\theta$ (or require that all HN factors of E have arguments strictly bigger than the $\mathrm{Arg}(E)$). For all described versions the Corollary 3.1.4 remains true.

[6] I thank the referee for the comments on those conditions.

3.3 Stable Framed Representations of Quivers

Let $\mathbf{k}$ be an algebraically closed field.

In the case of quivers without potential there is a well-known way (exploited by Nakajima and Reineke among others) to construct framed objects by adding a new vertex i_0 and d_i new arrows $i_0 \to i$ for each vertex $i \in I$ of the quiver Q. If we denote by W_i the vector space spanned by d_i arrows, then the framing functor Φ assigns to a representation $E = (E_i)_{i\in I}$ the vector space $\prod_{i\in I} \mathrm{Hom}(W_i, E_i)$. Let $\gamma = (\gamma^i) \in \mathbf{Z}^I_{\geq 0}$ be a dimension vector.

Then a framed representation of Q is given by a representation of the extended quiver $\widehat{Q}$ with the set of vertices $I \sqcup \{i_0\}$ of dimension $(\gamma^{i_0} = 1, (\gamma^i)_{i\in I})$, a collection of new d_i arrows $i_0 \to i, i \in I$.

Let us fix a central charge $Z : \mathbf{Z}^I \to \mathbf{C}$ and a ray $l := l_\theta = \mathbf{R}_{>0}e^{i\theta}, 0 < \theta \leq \pi$. Recall that we have the category $\mathcal{C}_l$ of semistable representations of Q with the central charge in l. A framed representation $E \in \mathcal{C}_l$ is stable framed if the following condition is satisfied (see [Re1]):

the representation E is semistable, and any subrepresentation E' which contains the images of all vector spaces $W_i, i \in I$ has a strictly smaller argument of the central charge.

There are many versions of the above criterion. For example, one can start with several additional vertices instead of just one. Also, one can restate the above criterion in terms of stable representations of the extended quiver $\widehat{Q}$. The later approach makes it clear why the notion of stable framed representation can be thought of as a generalization of the notion of a cyclic representation.

Remark 3.3.1. Let us set $d_1 = 1$ in the above notation. For the quiver Q_2 with one vertex and two loops there are no nontrivial stability conditions. Then stable framed objects is the same as left ideals of finite codimension in the path algebra of Q_2. The moduli space of stable framed objects is known as the *non-commutative Hilbert scheme of* $\mathbf{k}^2$.

4 Modules over COHA from Stable Framed Objects

4.1 Quiver Case

Let $\mathbf{k}$ be an algebraically closed field.

Let us fix a quiver Q with the set of vertices I as well as a central charge $Z : \mathbf{Z}^I \to \mathbf{C}$. We also fix a slope $0 < \theta \leq \pi$ and the corresponding ray $l = l_\theta = \mathbf{R}_{>0} \cdot e^{i\theta}$. In order to specify the framing we fix a collection $(d_i)_{i\in I}$ of non-negative integer numbers. An additional (framing) vertex is denoted by i_0. The corresponding extended quiver will be denoted by $\widehat{Q} := Q^{i_0}((d_i)_{i\in I})$.

Given a dimension vector $\gamma \in \mathbf{Z}^I_{\geq 0}$ we denote by $\mathsf{M}^{\mathrm{st}}_{\gamma,(d_i)_{i\in I}} := \mathsf{M}^{\mathrm{st,l}}_{\gamma,(d_i)_{i\in I}}$ the scheme of stable framed representations of dimension γ having $Z(\gamma) \in l$. We denote by $\mathsf{M}_{\gamma,(d_i)_{i\in I}} := \mathsf{M}^l_{\gamma,(d_i)_{i\in I}}$ the bigger space of framed representations (no stability conditions are imposed). The group $\mathsf{G}_\gamma = \prod_i GL(\gamma^i, \mathbf{k})$ acts freely on $\mathsf{M}^{\mathrm{st}}_{\gamma,(d_i)_{i\in I}}$. We denote by $V^l_{\gamma,(d_i)_{i\in I}} = V^\theta_{\gamma,(d_i)_{i\in I}}$ the graded vector space $H^\bullet_{\mathsf{G}_\gamma}(\mathsf{M}^{\mathrm{st}}_{\gamma,(d_i)_{i\in I}}) = H^\bullet(\mathsf{M}^{\mathrm{st}}_{\gamma,(d_i)_{i\in I}}/\mathsf{G}_\gamma)$.

Recall that with the ray $l = l_\theta$ we can associate COHA

$$\mathcal{H}_l = \oplus_{\gamma\in \mathbf{Z}^I_{\geq 0}, Z(\gamma)\in l} H^\bullet_{\mathsf{G}_\gamma}(\mathsf{M}^{ss}_\gamma).$$

Let us denote by $S := S_{\gamma_1,\gamma_2,\gamma_3,(d_i)_{i\in I}}$ the scheme of short exact sequences

$$0 \to E_1 \to E_2 \to E_3 \to 0$$

of such representations of the extended quiver $\widehat{Q}$ that $\dim(E_i) = \gamma_i \in \mathbf{Z}^{I\sqcup\{i_0\}}_{\geq 0}, i = 1, 2, 3$, E_2, E_3 stable framed, and the morphism $E_2 \to E_3$ is equal to the identity at the vertex i_0.

There is a projection $\pi_{13} : S \to \mathsf{M}_{\gamma_1} \times \mathsf{M}^{\mathrm{st}}_{\gamma_3,(d_i)_{i\in I}}$ which sends the short exact sequence $0 \to E_1 \to E_2 \to E_3 \to 0$ to the pair (E_1, E_3), where we treat E_1 as a representation of Q. Similarly we have a projection π_2 to E_2. Notice that the latter is a proper morphism of S to $\mathsf{M}^{\mathrm{st}}_{\gamma_2,(d_i)_{i\in I}}$. Since the automorphism group of the moduli space of stable framed objects is trivial, we see that the morphism $\pi_{2*}\pi^*_{13}$ gives rise to a map of cohomology groups

$$H^\bullet_{\mathsf{G}_{\gamma_1}}(\mathsf{M}_{\gamma_1}) \otimes H^\bullet(\mathsf{M}^{\mathrm{st}}_{\gamma_3,(d_i)_{i\in I}}) \to H^\bullet\left(\mathsf{M}^{\mathrm{st}}_{\gamma_2,(d_i)_{i\in I}}\right).$$

Proposition 4.1.1. *The above map gives rise to a (left) $\mathcal{H}_l$-modules structure on the vector space $V^l := V^l_{(d_i)_{i\in I}} = \oplus_\gamma V^l_{\gamma,(d_i)_{i\in I}}$.*

Proof. Similar to the proof of associativity of the product on COHA given in [KoSo5]. ■

Remark 4.1.2. The above considerations can be generalized to the case of quivers with potential.

Example 4.1.3. In the case of the quiver Q_2 (one vertex and two loops) and $d_1 = 1$ the moduli space $\mathsf{M}^{\mathrm{st}}_{\gamma,d_1}, \gamma \in \mathbf{Z}_{\geq 0}$ is the same as the moduli space of representations of the free algebra $\mathbf{k}\langle x_1, x_2\rangle$ of dimension γ which are cyclic. In other words, it is the moduli space of codimension γ ideals in the free algebra with two generators, i.e., it is the non-commutative Hilbert scheme. The above Proposition claims that it carries a structure of module over the COHA for Q_2 (which is the infinite Grassmann algebra). Explicit formulas for this module structure (and their generalization to the case of arbitrary number of loops) can be found in [Fra].

Consider COHA $\mathcal{H}$ of a quiver which has at least one vertex i_0 without loops. Then $\mathcal{H}$ is a module over the infinite Grassmann algebra (a.k.a free fermion algebra) $\Lambda^\bullet$. Indeed, consider i_0 as a quiver Q_0 (one vertex, no loops). We know that COHA of Q_0 is $\Lambda^\bullet$. Since it is a subalgebra of $\mathcal{H}$, it acts on $\mathcal{H}$ by left multiplication.

Let Q be a quiver with the set of vertices I. Let us fix a set of non-negative integers $d = (d^i)_{i\in I}$ (not all equal to zero)and the dimension vector $\gamma = (\gamma^i)_{i\in I}$. Then we have an extended quiver $\widehat{Q}$ with the set of vertices $I \sqcup i_0$ and d^i arrows from i_0 to $i \in I$. For a fixed central charge $Z : \mathbf{Z}^I \to \mathbf{C}$ the moduli space $\mathsf{M}^{\mathrm{st,l}}_{\gamma,d}$ of stable framed representations of Q of dimension γ such that $Z(\gamma) \in l$ is a non-empty smooth variety of pure dimension $\sum_{i\in I} d^i\gamma^i - \chi(\gamma,\gamma)$, where $\chi(\alpha,\beta)$ is the Euler–Ringel bilinear form of Q (see [EnRe], Proposition 3.6). Moreover it admits a projective morphism to the moduli space of polystables of fixed slope.

4.2 Representations of COHA in General Case

We will give a sketch of the construction.

Let $\mathcal{A}$ be "good' abelian subcategory in the $3CY$ category $\mathcal{C}$. We assume the conditions on the potential W which guarantee existence of COHA of $\mathcal{A}$ as well as moduli spaces of stable framed objects. Then considerations from the previous subsection can be generalized to this situation provided $\mathcal{A}$ satisfies some extra conditions, e.g., that classes $\mathrm{cl}(E)$ of objects of $\mathcal{A}$ belong to an additive monoid Γ_+ which is mapped to $\mathbf{Z}^n_{\geq 0}$ under the chosen identification $\Gamma \simeq \mathbf{Z}^n$.

Next, let us fix a ray $l = \mathbf{R}_{\geq 0}\cdot e^{i\theta}$ in the upper-half plane, and a stability function $Z : \Gamma \to \mathbf{C}$ such that $Z(\Gamma_+)$ belongs to the upper-half plane. Then we have the category $\mathcal{A}_l$ of semistables with the central charge in l. Let us fix the framing functor Φ. Then we can speak about framed and stable framed objects.

Recall that there is a notion of morphism of framed objects $(E_2,f_2) \to (E_3,f_3)$. An epimorphism $(E_2,f_2) \to (E_3,f_3)$ is a morphism in the category of framed objects which induces a homomorphism $H^0(\Phi(E_2)) \to H^0(\Phi(E_3))$ which sends f_2 to f_3 (see Sect. 3.1 for the notation).

Assume that E_2 and E_3 are semistable objects with central charges in the ray l. Then the kernel of the epimorphism $(E_2,f_2) \to (E_3,f_3)$ of framed objects does not have to be framed. Let us consider the stack $\mathcal{Z}_{\gamma_1,\gamma_2}$ of triples $(E_1,(E_2,f_2),(E_3,f_3))$ where:

(a) $\mathrm{cl}(E_1) = \gamma_1$, and $Z(\gamma_1) \in l$;
(b) (E_2,f_2) is stable framed, $\mathrm{cl}(E_2) = \gamma_1 + \gamma_2$, $Z(\mathrm{cl}(E_2) \in l$;
(c) (E_3,f_3) is stable framed, $\mathrm{cl}(E_3) = \gamma_2$, $Z(\mathrm{cl}(E_2) \in l$;
(d) there is a epimorphism of framed objects $(E_2,f_2) \to (E_3,f_3)$ such that it induces (in the category of semistable objects with the central charge in l) a short exact sequence

$$0 \to E_1 \to E_2 \to E_3 \to 0.$$

Recall that stable framed objects with fixed class $\gamma \in \Gamma_+$ form a scheme which we denote by $\mathcal{M}^{\mathrm{st,fr}}_{\gamma}$. Then we have natural projections $p_2 : \mathcal{Z}_{\gamma_1,\gamma_2} \to \mathcal{M}^{\mathrm{st,fr}}_{\gamma_1+\gamma_2}$ and $p_3 : \mathcal{Z}_{\gamma_1,\gamma_2} \to \mathcal{M}^{\mathrm{st,fr}}_{\gamma_2}$ which are morphisms of stacks. Furthermore, let $\mathcal{M}_\gamma$

denotes the moduli stack of objects of $\mathcal{A}_l$. Then we have the natural projection $p_1 : \mathcal{Z}_{\gamma_1,\gamma_2} \to \mathcal{M}_{\gamma_1}$.

We will assume that:

(i) if we consider the analog of the above situation with all $f_i = 0, i = 1, 2, 3$ (i.e., we work just in the abelian category $\mathcal{A}_l$), then the restriction of p_2 to $p_1^{-1}(E_1) \cap p_3^{-1}(E_3)$ is a morphism of smooth proper stacks;

(ii) in general, for fixed E_1 and (E_3, f_3) as above, the restriction of p_2 to $p_1^{-1}(E_1) \cap p_3^{-1}((E_3, f_3))$ is a morphism of smooth proper stacks.

By condition (i) COHA $\mathcal{H}_l$ of the category $\mathcal{A}_l$ is well-defined as an associative algebra. For that we use the critical version of the cohomology from [KoSo5] with trivial potential. Furthermore, repeating the construction from the previous subsection we obtain a structure of (left) $\mathcal{H}_l$-module on $V := V^l = \oplus_{\gamma\in\Gamma_+} H^\bullet(\mathcal{M}_\gamma^{\mathrm{st,fr}})$.

Remark 4.2.1. More generally, we can construct modules over COHA by considering the stack of objects whose Harder–Narasimhan filtration has consecutive factors with arguments of the central charge belonging to the interval $[\theta, \pi]$. If we have an exact short sequence

$$0 \to E_1 \to E_2 \to E_3 \to 0,$$

such that $\operatorname{Arg} Z(E_2) \in [\theta, \pi]$, then $\operatorname{Arg} Z(E_3)$ belongs to the same interval, while $\operatorname{Arg} Z(E_a) \in [0, \pi]$. Assume that all HN subquotients of E_2 have arguments in $[\theta, \pi]$.[7] Then by the above construction we obtain a representation of COHA in the cohomology of the stack of objects generated by semistables E such that $\operatorname{Arg} Z(E) \in [\theta, \pi]$.

Similarly, one can show that if V is a strict sector in the plane, then the graded "Cohomological Hall vector space" $\mathcal{H}(V)$ bounded from the left by a ray l is a module over the COHA $\mathcal{H}_l$ associated with the ray.

Furthermore, suppose that our abelian category $\mathcal{A}$ is a "good" subcategory of an ind-Artin $3CY$ category $\mathcal{C}$ endowed with orientation data. Let W be the potential for $\mathcal{C}$. It gives rise to the sheaf of vanishing cycles ϕ_W on the stack of objects of $\mathcal{C}$. Then the pull-backs of ϕ_W to the stack of objects of $\mathcal{A}$ and subsequently to $\mathcal{M}_\gamma$ and $\mathcal{M}_\gamma^{\mathrm{st,fr}}$ are well-defined. Then, similarly to [KoSo5] (and under the above assumption), the above construction (but this time with cohomology groups with coefficients in ϕ_W) gives rise to the module $V := V^l = \oplus_{\gamma\in\Gamma_+} H^\bullet(\mathcal{M}_\gamma^{\mathrm{st,fr}}, \phi_W)$ over the COHA $\mathcal{H}_l$ of $\mathcal{A}_l$. The details will be explained elsewhere.

[7] I thank the referee for pointing out on this condition.

4.3 Hecke Operators Associated with Simple Objects

In the classical Nakajima construction of the infinite Heisenberg algebra (see [Nak2]) one considers pairs of ideal sheaves (J_2, J_3) on a surface S such that $J_2 \subset J_3$ and $\mathrm{Supp}(J_3/J_2) = \{x\}$, where x is a fixed point. Then one has an epimorphism $\mathcal{O}_S/J_2 \to \mathcal{O}_S/J_3$. Let us compare this observation with the above construction of modules over COHA. We see that a fixation of K-theory classes $\gamma_i, i = 1, 2$ for a pair of objects (E_2, E_3) along with an epimorphism $E_2 \to E_3$ corresponds in the Nakajima's construction to the fixation of $n_i, i = 2, 3$ such that $J_i \in \mathrm{Hilb}_{n_i}(S)$ and to the above-mentioned epimorphism of the quotient sheaves.

In the construction of the module structure on the cohomology of stable framed objects we used the pushforward map associated with the projection to the middle term in the moduli space of short exact sequences

$$0 \to E_1 \to E_2 \to E_3 \to 0,$$

where E_2, E_3 are stable framed (we omit here $f_i, i = 2, 3$ from the notation). As a result, our construction gives rise to the "raising degree" operators $\mathcal{H}_{\gamma_1} \otimes V_{\gamma_2} \to V_{\gamma_1+\gamma_2}$ for the COHA action $\mathcal{H} \otimes V \to V$. There are no "lowering degree" operators, which would correspond to the projection to the term $E_3 = E_2/E_1$. The reason is similar to the one in the Nakajima's construction: such a projection is not proper.

Originally Nakajima solved the problem by considering points $x \in S$ which belong to a compact subset in S. We can use this idea and consider short exact sequences as above, where E_1 is a *simple* object which runs through a compact (in analytic topology) subset in the moduli scheme of simple objects of our abelian category $\mathcal{A}$.

Let us illustrate the construction in the case of quivers without potential and trivial stability condition. In that case stable framed objects are cyclic modules over the path algebra of the quiver. Then we should prove that there are sufficiently many cyclic modules with the fixed simple submodule and fixed cyclic quotient. This is guaranteed by the following result.

Proposition 4.3.1. *Let A be an associative algebra, $(M_2, v_2), (M_3, v_3)$ be A-modules with marked elements $v_i \in M_i, i = 2, 3$ such that v_3 is a cyclic vector for M_3. Let $f : M_2 \to M_3$ be an epimorphism of A-modules such that $f(v_2) = v_3$ and such that $W = \mathrm{Ker}(f)$ is a simple A-module. Suppose that the extension*

$$0 \to W \to M_2 \to M_3 \to 0$$

is non-trivial. Then v_2 is a cyclic vector for M_2.

Proof. Let $M_2' \subset M_2$ be the A-submodule generated by v_2. If $M_2' = M_2$, then we are done. Otherwise we have a non-trivial epimorphism $g : W \to M_2/M_2'$ of A-modules. Its kernel is a submodule of W. It must be trivial, since W is simple. Hence g is an isomorphism. Then the submodules W and M_2' determine the direct sum decomposition $M_2 = W \oplus M_2'$, where $M_2' \simeq M_3$. Hence the extension $0 \to W \to M_2 \to M_3 \to 0$ is trivial. This contradiction shows that v_2 is a cyclic vector. ■

Corollary 4.3.2. *For fixed W, M_3 the stack of cyclic modules M_2 which are middle terms in the above short exact sequence is a smooth projective scheme isomorphic to the projective space* $\mathbf{P}(\mathrm{Ext}^1(M_3, W))$.

Proof. Follows from the Proposition. ■

Remark 4.3.3. Similar result holds in case when M_3 is stable framed and S is simple.

Let now $\mathcal{M}^{\mathrm{simp}} := \mathcal{M}^{\mathrm{simp}}_{\mathcal{A}}$ be the moduli space of simple objects in the heart $\mathcal{A}$ of the "good" t-structure of an ind-Artin $3CY$ category $\mathcal{C}$ endowed with orientation data. Then $\mathcal{M}^{\mathrm{simp}}_{\mathcal{A}}$ is a smooth separated scheme. Let $H^{\bullet}_c(\mathcal{M}^{\mathrm{simp}})$ denotes compactly supported cohomology. As before we have projections π_1, π_2, π_3. In particular, π_1 is the map to the moduli space $\mathcal{M}^{\mathrm{simp}}$ of simple objects and π_3 is the map to the moduli space $\mathcal{M}^{\mathrm{st}}$ of stable framed objects correspondingly. Then the composition $\pi_{3*} \circ (\pi_1^* \otimes \pi_2^*)$ defines a collection of operations on $H^{\bullet}(\mathcal{M}^{\mathrm{st}})$ parametrized by the elements of $H^{\bullet}_c(\mathcal{M}^{\mathrm{simp}})$. The above Proposition (or rather its analog for non-trivial stability condition) ensures that the operations are well-defined. Differently from the action of COHA defined in the previous subsection, these operations *decrease* the degree $\gamma \in \Gamma$.

Remark 4.3.4. Let us recall that for any $i \in \pm\mathbf{Z}_{>0}$ Nakajima defines an operator $P[i]$ which corresponds to the i-th generator of the infinite Heisenberg algebra. In the above discussion the operator $P[i]$ corresponds to the direct sum $iS := S \oplus S \oplus \ldots \oplus S$ of $\pm i > 0$ of copies of the simple object S.

Using the above construction one can extend a representation of COHA to a representation of a bigger algebra, which we call "full COHA" (or double of COHA). We do not know yet how to define this algebra intrinsically. Our approach is similar to the Nakajima's construction of the infinite Heisenberg algebra from two representations of the symmetric algebra: one is given by creation operators and another one is given by annihilation operators. Commuting creation and annihilation representations *in the representation space* Nakajima recovers the infinite Heisenberg algebra. One can also compare the above construction with the one in [Re1].

Remark 4.3.5. As we already mentioned, differently from the case of constructible Hall algebras, we do not know a compatible comultiplication on COHA. This would help to define full COHA by means of the Drinfeld double construction. Having in mind that in the case of quivers without potential COHA is a shuffle algebra, one can hope for explicit formulas similar to those in [Neg] (see also [Dav2]).

4.4 Full COHA: An Example

It is well known that one can obtain finite-dimensional representations of a quantized enveloping algebra of finite-dimensional semisimple Lie algebra using framed stable representations of quivers and constructible Hall algebra (see, e.g., [Re1] where this idea which goes back to Nakajima was implemented for representations of the "positive" part of a quantum group).

Let us recall how one can recover the quantized enveloping algebra $U_q(\mathrm{sl}(2))$ by combining two representations of the Hall algebra for quiver A_1.

Recall that if for a quiver Q we take the stability function $\Theta = 0$, then every finite-dimensional representation of Q is semistable. The moduli space of stable framed representations admits in this case a simple description in terms of Grassmannians (see, e.g., [Re1], Proposition 3.9).

For the quiver A_1 (one vertex i_1 and no arrows) the framing consists of a new vertex i_0 and d arrows $i_0 \to i_1$. The stability function is trivial automatically, and one can easily see that for each dimension vector $\gamma \in \mathbf{Z}_{\geq 0}$ the moduli space $\mathcal{M}_{\gamma,d} = \mathsf{M}_{\gamma}^{\theta=0,\mathrm{st}}$ of framed stable representations of dimension γ is isomorphic to the Grassmannian $\mathrm{Gr}(d-\gamma,d) \simeq \mathrm{Gr}(\gamma,d)$. Hence it is non-empty for $\gamma \leq d$ only. Let us denote by $\mathrm{Gr}(d)$ the "full Grassmannian" consisting of vector subspaces of $\mathbf{C}^d$ of all dimensions (this space is disconnected). Then the moduli space of d-framed semistable representations of A_1 is $\mathrm{Gr}(d)$.

The space of $\mathrm{GL}(d)$-invariant functions with finite support $\mathrm{Fun}^{\mathrm{GL}(d)}(\mathrm{Gr}(d))$ is a module over the constructible Hall algebra of A_1. The constructible Hall algebra of the quiver A_1 is the polynomial algebra with one generator $z := \mathbf{1}_{\mathbf{C}}$, where the generator z corresponds to the characteristic function $\mathbf{1}_1$ of M_1 in the stack $\mathcal{M} = \sqcup_{\gamma \geq 0} \mathsf{M}_{\gamma}$. Indeed the Hall product gives an isomorphism of the constructible Hall algebra with the polynomial ring $\mathbf{C}[z]$. In each $\mathrm{Gr}(k,d)$ we have only one $\mathrm{GL}(d)$-orbit of the standard coordinate vector subspace $\mathbf{C}^k \subset \mathbf{C}^d$. Let us denote by $v_k, 0 \leq k \leq d$ the characteristic function of the corresponding $\mathrm{GL}(d)$-orbit.

Let us consider the "minus" Hecke correspondence given by pairs $(V_{k-1} \subset V_k)$ with 1-dimensional factor V_1 and project to V_{k-1}. Equivalently, we consider the projection to the first terms from the set of short exact sequences

$$0 \to V_{k-1} \to V_k \to V_1 \to 0.$$

Using the pull-back/pushforward construction discussed previously, we obtain a representation of the algebra $\mathbf{C}[z]$ given by $\rho_-(z)v_k = \frac{q^k - q^{-k}}{q - q^{-1}} v_{k-1}, 1 \leq k \leq d$, and $\rho_-(z)v_0 = 0$, where the factor comes from the normalization of the cocycle $c(M,N)$ above as $q^{\chi(M,N)}$. The Euler–Ringel form χ on the pair of representations E of dimension a and F of dimension b is given by $\chi(E,F) = ab$. Similarly, consider the "plus" Hecke correspondence $(V_k \subset V_{k+1})$ and project to V_{k+1}. Then we get a representation of $\mathbf{C}[z]$ in $\mathcal{F}_n$ given by

$$\rho_+(z)v_k = \frac{q^{k+1} - q^{-k-1}}{q - q^{-1}} v_{k+1}, 0 \le k \le d-1, \rho_+(z)v_d = 0.$$

Combining ρ_- and ρ_+ together we obtain the standard d-dimensional representation of the quantized enveloping algebra $U_q(\mathrm{sl}(2))$ where the "positive" generator E is represented by $\rho_-(z)$ while the "negative" generator F is represented by $\rho_+(z)$. Then the commutator $[E, F]$ maps v_k to $\frac{q^{2k}-q^{2k}}{q-q^{-1}} v_k$. From this formula one can recover the action of the Cartan generators K, K^{-1}.

Let us apply similar considerations in the case of COHA of the quiver A_1. Recall, in this case COHA is isomorphic to the algebra $\Lambda^\bullet = \Lambda^\bullet(\xi_1, \xi_2, \ldots), \deg \xi_i = 2i-1, i \ge 1$. Since for the trivial stability function the COHA associated with the ray $\theta = 0$ coincides with whole COHA, we obtain a representation of the infinite Grassmann algebra $\Lambda^\bullet$ in the finite-dimensional vector space $V := H^\bullet(\mathrm{Gr}(d)) = \oplus_{0 \le k \le d} H^\bullet(\mathrm{Gr}(k, d))$. One can write down explicitly the action of the generators of $\Lambda^\bullet$ on the cohomology classes of Schubert cells.

Following the general definition, we consider the moduli stack of short exact sequences

$$0 \to E_1 \to E_2 \to E_3 \to 0,$$

where E_2 and E_3 are stable framed of the same slope, and E_1 is a representation without framing. First we consider the representation of COHA coming from the projection to E_2.

In order to write down the corresponding representation explicitly let us choose a subspace in each $H^\bullet(\mathrm{Gr}(k, d)), 0 \le k \le d$ spanned by the cohomology classes corresponding to $(\mathbf{C}^*)^d$-fixed points. We denote this basis by $e_j := \mathbf{1}_{\mathbf{C}_{j_1,\ldots j_k}}$ (recall that the fixed points correspond to coordinate subspaces $\mathbf{C}_{j_1,\ldots j_k} \subset \mathbf{C}^d$ spanned by the standard basis vectors $f_{j_1}, \ldots, f_{j_k}, j_1 < j_2 < \ldots < j_k$). We can identify the graded vector space V with the quotient $\Lambda^\bullet(\xi_1, \ldots, \xi_d)/I_d$, where I_d is the ideal generated by $\xi_i, i \ge d+1$. Then the pull-back/pushforward construction gives us the representation of $\Lambda^\bullet$ in V by "creation" operators: $a_n^* : e_j \mapsto \xi_n \wedge e_j$ (see also [Fra]).

Second, we consider the representation of COHA coming from the projection to E_3. This gives a representation of $\Lambda^\bullet$ on V by "annihilation" operators $a_n : e_j \mapsto i_{\xi_n}(e_j)$, where i_{ξ_n} is the contraction operator which deletes the variable ξ_n from the monomial e_j.

Then, similarly to the consideration with the constructible Hall algebra, we combine both representations of COHA into a single one. In this way recovers the representation of the Lie algebras D_{d+1} (see [Xi]). This leads to the following conjecture:

Conjecture 4.4.1. Full COHA for the quiver which has one vertex and m arrows is isomorphic for even m to the infinite Clifford algebra Cl_c with generators $\xi_n^\pm, n \in 2\mathbf{Z}+1$ and the central element c, subject to the anticommuting relations between ξ_n^+ (resp. ξ_n^-) as well as the relation $\xi_n^+\xi_m^- + \xi_m^-\xi_n^+ = \delta_{nm}c$.

In the case of odd m full COHA is isomorphic to the infinite (graded) Heisenberg algebra (in the above formulas change anticommuting brackets by the commuting ones).

In the case of finite-dimensional representations we have $c \mapsto 0$ and two representations of the infinite Grassmann algebra, which are combined in the representations of the orthogonal Lie algebra as explained above. Notice that Cl_c is the Clifford algebra associated with the positive part of the affine Lie algebra sl(2). This might explain the relation of the full COHA to the quiver A_1.

5 Some Representations of COHA Motivated by Physics and Knot Theory

In this section we are going to describe some interesting classes of representations of COHA. Details of the constructions will appear elsewhere. The reader can consider this section as a collection of speculations.

5.1 Fukaya Categories of Conic Bundles and Gauge Theories from Class $\mathcal{S}$

We are going to illustrate the idea in the case of $SL(2)$ Hitchin integrable systems. Our motivation is the general conjecture (F.1) from the Introduction of [ChDiManMoSo]. In this particular case it admits a very precise interpretation. Namely, with a point of the universal cover of the base of Hitchin system on a complex curve, say, C one can associate a compact Fukaya category endowed with a stability structure ("compact" means that it is generated by local systems supported on compact Lagrangian submanifolds). It is the Fukaya category of a non-compact Calabi–Yau three-fold X described by the corresponding spectral curve (see [KoSo8] for a more general framework). The compact Fukaya category is endowed with the natural t-structure generated by SLAGs which are 3-dimensional Lagrangian spheres. The central charge of the corresponding stability condition is given by the period map of the Liouville form on T^*C restricted to the spectral curve. According to the general theory developed in Sect. 8 of [KoSo1] categories generated by spherical collections are in one-to-one correspondence with pairs (Q, W), i.e., quivers with potential. Hence we can speak about corresponding COHA and its representations in the cohomology of the moduli spaces of stable framed objects of the category $\mathrm{Crit}(W)$. This would give an interesting class of representations of the BPS algebra of the corresponding gauge theory from class $\mathcal{S}$.

Recall that surface defects in physics correspond to points of the curve C. In terms of the corresponding non-compact Calabi–Yau three-fold they are complex 2-dimensional submanifolds of X. Consider the moduli space of SLAGs with the boundary which belongs to such a submanifold S. The corresponding category $\mathcal{F}(X, S)$ can be thought of as a version of Fukaya–Seidel category of thimbles (see [Se]) with the (analog of the) potential being the natural map $X \to C$.

Furthermore, the operation of connected Lagrangian sum plays a role of an extension in the compact Fukaya category $\mathcal{F}(X)$. This operation underlies the product structure on the COHA $\mathcal{H}^{Q,W}$.

Let us observe that there is an operation of taking the connected Lagrangian sum of a Lagrangian submanifold without boundary and the one with the boundary on S. Mimicking the definition of the product on COHA with the "moduli space of Lagrangian connected sums" instead of the subvariety $\mathsf{M}_{\gamma_1,\gamma_2} \subset \mathsf{M}_{\gamma_1+\gamma_2}$ (see [KoSo2], Sect. 2), one obtains the $\mathcal{H}^{Q,W}$-module structure on the cohomology of the moduli space of SLAGs with the boundary on S.[8]

5.2 Resolved Conifold and Quivers

Let $X = \mathrm{tot}(\mathcal{O}(-1) \oplus \mathcal{O}(-1))$ be the resolved conifold. We denote the zero section of the corresponding vector bundle by $C_0 \simeq \mathbf{P}^1$. Let us fix a point $p_0 := 0 \in C_0$.

Let $\mathcal{A}$ be the abelian category of *perverse coherent sheaves* on X topologically supported on C_0 (see, e.g., [NagNak, Tod] for descriptions convenient for our purposes; in [NagNak] our category $\mathcal{A}$ was denoted by $Per_c(X/Y)$, where $X \to Y$ is the crepant resolution of the conifold singularity $Y = \{xy - zw = 0\}$).

It is known (see, e.g., [NagNak]) that $\mathcal{A}$ is equivalent to the abelian category $\mathrm{Crit}(W)$ associated with the pair (Q, W), where Q is a quiver with two vertices i_1, i_2 two arrows $a_1, a_2 : i_1 \to i_2$, two arrows $b_1, b_2 : i_2 \to i_1$ and "Klebanov-Witten potential" $W = a_1b_1a_2b_2 - a_1b_2a_2b_1$. In particular, for any $\gamma = (\gamma^1, \gamma^2) \in \mathbf{Z}^2_{\geq 0}$ the stack of objects of $\mathrm{Crit}(W)$ of dimension γ is equivalent to the stack of such representations of Q of dimension γ in coordinate vector spaces, which belong to the critical locus of the function $\mathrm{Tr}(W)$.

We recall that the category of perverse coherent sheaves carries a family of geometrically defined *weak stability conditions* (see, e.g., [Tod]). In the case of the category $\mathrm{Crit}(W)$ there is a class of stability conditions associated with the slope function.

Equivalence of these two categories gives rise to the "chamber" structure of the space of stability conditions on $\mathcal{A}$ described in [NagNak]: some of the (infinitely many) chambers correspond to the quiver-type stability conditions, while "at infinity" we have chambers of geometric origin corresponding to different choices of the weak stability condition.

[8] Alternatively, following Paul Seidel, one can consider the double cover of the Calabi–Yau threefold branched along the divisor given by the complex surface. Then Lagrangian submanifolds with boundary lift to closed ones in the branched cover. One can form an equivariant Lagrangian connected sum, and then interpret it as an operation on the original Lagrangian submanifolds with boundary.

One has a similar story when the framing is taken into account. Then one deals with framed perverse coherent sheaves and framed representations of (Q, W) (i.e., critical points of the function $\mathrm{Tr}(W)$ considered as a function on the space of representations of the extended quiver $\widehat{Q}$ obtained from Q by adding an extra vertex i_0 and an arrow $i_0 \to i_1$.)

In the "quiver chamber," we can (after a choice of a stability condition on $\mathcal{A}$ which belongs to the above class) speak about the moduli space of stable framed objects of the fixed slope. For a given $(l, n) \in \mathbf{Z}_{\geq 0} \times \mathbf{Z}$ and a choice of certain stability condition on the category $Rep(Q)$ of finite-dimensional representations of Q, and a choice of an angle θ (which depends on (l, n)), the space of stable framed representations of (Q, W) with the slope θ becomes isomorphic to the moduli space of Pandharipande–Thomas stable pairs $P(l, n)$ (see [NagNak]). There is no single θ which serves all (l, n).

It follows from the previous section that:

Proposition 5.2.1. *For a choice of stability conditions in the "quiver chamber," COHA $\mathcal{H}^{(Q,W)}$ acts on the cohomology of the moduli space of stable framed representations of (Q, W) having fixed slope.*

Let us observe that if we have a morphism $f : E_2 \to E_3$ of PT stable pairs which is surjective in degree zero (i.e., on the sheaves supported on C_0), then $\mathrm{Ker}(f)$ is a coherent sheaf scheme-theoretically supported on C_0. Passing to the cohomology groups we reformulate the above Proposition by saying that COHA of the corresponding category acts on the cohomology of the moduli space of PT stable pairs. We expect the same result to hold in "geometric chambers," were one uses weak stability conditions.

5.3 *Vertically Framed Sheaves and Algebraic Knots*

Motivated by [GuSchVa] and many other physics papers on the relation between knot theory and BPS states for the resolved conifold, one can hope for an application of the representation theory of COHA of the resolved conifold to knot invariants. Among mathematical motivations we can mention the main conjecture from [ORS], its reformulation in [DiHuSo] and its proof in [Mau1] (in the "unrefined" form). In this subsection we discuss appropriate moduli spaces following [DiHuSo] and speculate about corresponding representation of COHA.

Let K be an algebraic knot or link which is obtained by Milnor construction, i.e., via the intersection of a plane singular curve C_K with the S^3-boundary of a small ball around the singularity. If we would like to incorporate algebraic knots in the story, we should add to the story coherent sheaves on X supported on the curve C_K placed in the fiber of the projection $X \to C_0 = \mathbf{P}^1$.

More precisely we consider coherent sheaves which are "vertically framed" along C_K (see details [DiHuSo]). Stable vertically framed coherent sheaves provide a natural generalization of PT stable pairs from [PT1] see also [**?**] for some physics arguments.

Let us recall some general definitions and details following [DiHuSo].

Let $X = \mathrm{tot}(\mathcal{O}(-1) \oplus \mathcal{O}(-1))$ be the resolved conifold, C a planar complex algebraic curve with the only singular point p.

The abelian category of C-framed perverse coherent sheaves is a full subcategory $\mathcal{A}^C \subset D^b(\mathrm{Coh}(X))$. Roughly speaking, $\mathcal{A}^C$ consists of complexes E of coherent sheaves on X such that the cohomology sheaves $H^i(E)$ are non-trivial for $i \in \{0, -1\}$ only, and those cohomology sheaves are topologically supported one the union $C \cup C_0 = \mathbf{P}^1$ (see, loc. cit., Sect. 2.2 and below for more precise description). The category $\mathcal{A}^C$ is closed under extensions, and it is a full subcategory of the category of perverse coherent sheaves $\mathcal{A} \subset D^b(X)$. After fixing Kähler class ω on the compactification $\overline{X}$ defined in [DiHuSo], one defines a family of weak stability conditions on $\mathcal{A}^C$ associated with an explicitly given family of slope functions $\mu_b := \mu_{(\omega, b\omega)}$ described in the loc. cit. Then one can speak about C-framed (semi)stable sheaves, meaning weakly (semi)stable objects of $\mathcal{A}^C$ with respect to the slope function μ_b.[9]

For "very negative" value of b the moduli space $\mathcal{P}_b^s(X, C, r, n)$ of C-framed μ_b-stable objects E with $ch(E) = (-1, 0, [C] + r[C_0], n)$ is isomorphic to the moduli space of stable framed pairs on X in the sense of Pandharipande and Thomas which are C-framed. If we move the value of b from $b = -\infty$ to a small positive number (which depend on r) the above moduli space of μ_b-stable objects experiences finitely many wall-crossings. One of the main results of [DiHuSo] is a theorem which relates the moduli space of μ_b-stable objects of $\mathcal{A}^C$ for small $b > 0$ with the punctual Hilbert schemes from [ORS]. This relates the DT-invariants of the category of C-framed stable sheaves with HOMFLY polynomials of algebraic knots. The moduli space $\mathcal{P}_b^{ss}(X, C, r, n)$ of μ_b semistable objects is a $\mathbf{C}^*$-gerbe over $\mathcal{P}_b^s(X, C, r, n)$.

Let us fix $(r, n) \in \mathbf{Z}_{\geq 0} \times \mathbf{Z}$ and consider the full subcategory $\mathcal{A}_{r,n}^C \subset \mathcal{A}^C$ consisting of objects E such that $ch(E) = (-1, 0, [C] + r[C_0], n)$. Let E_1 be a pure dimension one sheaf on X supported on C_0 (hence it belongs to $\mathcal{A}^C$ as well), and let $E_3 \in \mathrm{Ob}(\mathcal{A}_{r_3,n_3}^C)$. Then we see that the middle term E_2 of an extension in $\mathcal{A}^C$

$$0 \to E_1 \to E_2 \to E_3 \to 0$$

belongs to $\mathcal{A}_{r_2,n_2}^C$ for some r_2, n_2.

Let $\mathcal{M}^C := \cup_{r,n} \mathcal{M}_{r,n}^C$ be the moduli space (stack) of the objects E which belong the category $\mathcal{A}_{r,n}^C$ for some r, n. Let $\mathcal{M}^{C_0}$ be the moduli space (stack) of objects of the category $\mathrm{Coh}_{C_0}(X)$ of coherent sheaves on X supported on the zero section $C_0 = \mathbf{P}^1$. Let $\mathcal{N}$ be the moduli space (stack) of short exact sequences as above.

We have the following projections: $\pi_{13} : \mathcal{N} \to \mathcal{M}^C \times \mathcal{M}^{c_0}, (E_1, E_2, E_3) \mapsto (E_1, E_3)$ and $\pi_2 : \mathcal{N} \to \mathcal{M}^C, (E_1, E_2, E_3) \mapsto E_2$.

[9] In [DiHuSo] the authors considered stable vertically framed sheaves on the compactification $\overline{X}$. The corresponding moduli spaces were projective. Considerations with non-compact submanifold X give rise to quasi-projective moduli spaces. We ignore these technicalities here.

Then we can apply the same procedure as for framed representations of quivers using the composition $\pi_{2*}\pi_{13}^*$. Then, e.g., in the case of COHA it gives us the module structure $H^\bullet(\mathcal{M}^{C_0})\otimes H^\bullet(\mathcal{M}^C) \to H^\bullet(\mathcal{M}^C)$ over the COHA of the category $\mathrm{Coh}_{C_0}(X)$, where by $H^\bullet$ we denote an appropriate cohomology theory.

Now we can use the weak stability condition defined by the slope function μ_b. More precisely, let us choose a stability parameter b satisfying the condition (3.1) of Lemma 3.1 from [DiHuSo] and consider μ_b-semistable objects E of $\mathcal{A}^C$ such that $ch(E) = (-1, 0, [C] + r[C_0], n)$, where $(r, n) \in \mathbf{Z}_{\geq 0} \times \mathbf{Z}$ is fixed. Then we can repeat the above definition but this time in the exact sequence

$$0 \to E_1 \to E_2 \to E_3 \to 0$$

we will assume that E_2 and E_3 are weakly semistable objects with respect to μ_b, and E_1, as before, is an arbitrary coherent sheaf on X supported on C_0.

There is an explicit description of μ_b-semistable and μ_b-stable objects of $\mathcal{A}^C_{r,n}$ for sufficiently small positive b given in [DiHuSo], Sect. 3. For example, a μ_b-stable object E fits into an exact short sequence

$$0 \to E_C \to E \to \mathcal{O}_{C_0}(-1)^r \to 0,$$

where $E_C = (\mathcal{O}_X \to F_C)$ is a stable pair on X in the sense of Pandharipande and Thomas, with the sheaf F_C scheme theoretically supported on C (and satisfying some non-degeneracy conditions, see [DiHuSo], Proposition 3.3 for the details). Similarly, any μ_b-semistable object fits into an exact sequence where instead of $\mathcal{O}_{C_0}(-1)^r$ one has a sheaf G topologically supported on C_0 (and $ch_2(G) = r[C_0]$) which is a direct image (under the embedding $i : C_0 \to X$) of the vector bundle $\oplus_{1\leq j\leq m}\mathcal{O}(a_j)^{r_j}$ with $a_1 > \ldots > a_m \geq -1$. The Harder–Narasimhan filtration of G (with respect to the ω-slope defined by $\chi(G)/r$) therefore has consecutive factors with slopes a_j/r.

Based on the above considerations one can hope that C-framed stable sheaves play a role similar to the one played by stable framed objects in the abelian categories. In particular, cohomology groups of the moduli spaces of C-framed stable sheaves should give rise to representations of COHA of X. It is not clear at this time how far this idea can be developed. In fact computations made by E. Diaconescu show that if in the short exact sequence $0 \to F \to E_1 \to E_2 \to 0$ the terms E_1, E_2 are C-framed stable, then F is isomorphic to $\mathcal{O}(-2)^n$. It seems plausible that in order to obtain interesting representations of the full COHA, one should also include in considerations short exact sequences of the type $0 \to E_1 \to E_2 \to F \to 0$, where E_1, E_2 are C-framed stable. This should lead to the representation of the full COHA in the way discussed previously. We expect that in this way we obtain affine sl(2).

Remark 5.3.1. The above story with C-framed sheaves is related to algebraic knots. As for more general knots, one can hope that the following picture can be made mathematically precise.

For any non-compact real-analytic Lagrangian submanifold $L \subset X$ with "good behavior at infinity" there should be a well-defined stack $\mathrm{Coh}_{\leq 1}(X, L)$ of real-analytic sheaves on X (considered as a real-analytic manifold) with the following properties:

(a) Every $F \in \mathrm{Coh}_{\leq 1}(X, L)$ has topological support, which is an immersed 2-dimensional real-analytic submanifold of X. Moreover, the support without boundary is an immersed non-compact complex analytic curve. The restriction of F to the complement of the boundary is a coherent sheaf on the corresponding complex manifold.
(b) The boundary of the support of each $F \in \mathrm{Coh}_{\leq 1}(X, L)$ belongs to L.
(c) The stack $\mathrm{Coh}_{\leq 1}(X, L)$ is a countable union of real-analytic stacks of finite type. It is naturally the stack of objects of the abelian category of real-analytic sheaves on X satisfying conditions (a) and (b).

In particular, sheaves F with pure support are those for which the support is an immersed "bordered Riemann surfaces" in the sense of [KatzLiu].

One can hope that despite of the analytic nature of objects, there is a theory of stability structures for this category, as well as the notion of stable framed object.

Notice that we can consider extensions $0 \to F \to E_1 \to E_2 \to 0$, where E_1, E_2 are objects of $\mathrm{Coh}_{\leq 1}(X, L)$, while F is the usual coherent sheaf on X with support on $C_0 = \mathbf{P}^1$. We expect that this operation leads to the action of COHA on the cohomology of framed stable objects in $\mathrm{Coh}_{\leq 1}(X, L)$, similarly to the case of C-framed stable sheaves.

Finally, if the above discussion about representations of COHA of the resolved conifold makes sense, then one can hope that it is related to the representation theory of DAHA discussed in [GorORS].

Acknowledgements I thank Alexander Braverman, Emanuel Diaconescu, Sergei Gukov, Nigel Hitchin, Mikhail Kapranov, Gabriel Kerr, Melissa Liu, Hiraku Nakajima, Andy Neitzke, Alexei Oblomkov, Vivek Schende, Olivier Schiffmann, and Xinli Xiao for stimulating discussions and correspondences. I also thank the anonymous referee for multiple comments leading to the improvement of the paper. I am especially grateful to Maxim Kontsevich for numerous discussions on the subject of this paper and related projects. I also thank IHES for excellent research conditions. This work is partially supported by an NSF grant.

References

[BLM] A. Beilinson, G. Lusztig, R. Macpherson, A geometric setting for the quantum deformation of GL_n. Duke Math. J. **61**, 2 (1990)

[BuJoMe] V. Bussi, D. Joyce, S. Meinhardt, On motivic vanishing cycles of critical loci. arXiv:1305.6428

[ChDiManMoSo] W. Chuang, D. Diaconescu, J. Manschot, G. Moore, Y. Soibelman, Geometric engineering of (framed) BPS states. arXiv:1301.3065

[Dav1] B. Davison, Invariance of orientation data for ind-constructible Calabi-Yau A_∞-categories under derived equivalence. arXiv:1006.5475
[Dav2] B. Davison, The critical CoHA of a self dual quiver with potential. arXiv:1311.7172
[Dav3] B. Davison, Purity of critical cohomology and Kac's conjecture. arXiv:1311.6989
[DiHuSo] D.-E. Diaconescu, Z. Hua, Y. Soibelman, HOMFLY polynomials, stable pairs and Donaldson-Thomas invariants. arXiv:1202.4651
[DiSVa] D.-E. Diaconescu, V. Shende, C. Vafa, Large N duality, Lagrangian cycles, and algebraic knots, arXiv:1111.6533
[DyKap] T. Dyckerhoff, M. Kapranov, Higher Segal spaces I. arXiv:1212.3563
[Ef] A. Efimov, Cohomological Hall algebra of a symmetric quiver. arXiv:1103.2736
[EnRe] J. Engel, M. Reineke, Smooth models of quiver moduli. arXiv:0706.43.06
[Fra] H. Franzen, On cohomology rings of non-commutative Hilbert schemes and CoHa-module. arXiv:1312.1499
[Ga1] D. Gaiotto, $\mathcal{N} = 2$ dualities. arXiv:0904.2715
[GaMoNe-2] D. Gaiotto, G. Moore, D. Neitzke, Wall-crossing, Hitchin systems and WKB approximations. arXiv:0907.3987
[GorORS] E. Gorsky, A. Oblomkov, J. Rasmussen, V. Shende, Torus knots and the rational DAHA, arXiv:1207.4523
[GuSto] S. Gukov, M. Stosic, Homological algebra of knots and BPS states. arXiv:1112.0030
[GuSchVa] S. Gukov, A. Schwarz, C. Vafa, Khovanov-Rozansky homology and topological strings. arXiv:hep-th/0412243
[H] D. Huybrechts, Introduction to stability conditions. arXiv:1111.1745
[HaMo1] J. Harvey, G. Moore, Algebras, BPS states, and strings. arXiv:hep-th/9510182
[HaMo2] J. Harvey, G. Moore, On the algebras of BPS states. arXiv:hep-th/9609017
[Hu] Z. Hua, Spin structure on moduli space of sheaves on Calabi-Yau threefold. arXiv:1212.3790
[KaKoPa] L. Katzarkov, M. Kontsevich, T. Pantev, Hodge theoretic aspects of mirror symmetry. arXiv:0806.0107
[KatzLiu] S. Katz, C. Liu, Enumerative geometry of stable maps with Lagrangian boundary conditions and multiple covers of the disc. arXiv:math/0103074
[Ko1] M. Kontsevich, Holonomic D-modules and positive characteristic. arXiv:1010.2908
[KoSo1] M. Kontsevich, Y. Soibelman, Stability structures, motivic Donaldson-Thomas invariants and cluster transformations. arXiv:0811.2435
[KoSo2] M. Kontsevich, Y. Soibelman, Affine structures and non-archimedean analytic spaces. math.AG/0406564
[KoSo3] M. Kontsevich, Y. Soibelman, Notes on A-infinity algebras, A-infinity categories and non-commutative geometry. I. math.RA/0606241
[KoSo4] M. Kontsevich, Y. Soibelman, Deformations of algebras over operads and Deligne's conjecture. arXiv:math/0001151
[KoSo5] M. Kontsevich, Y. Soibelman, Cohomological Hall algebra, exponential Hodge structures and motivic Donaldson-Thomas invariants. arXiv:1006.2706
[KoSo6] M. Kontsevich, Y. Soibelman, Homological mirror symmetry and torus fibrations. arXiv:math/0011041
[KoSo7] M. Kontsevich, Y. Soibelman, Lectures on motivic Donaldson-Thomas invariants and wall-crossing formulas (preprint, 2010)
[KoSo8] M. Kontsevich, Y. Soibelman, Wall-crossing structures in Donaldson-Thomas invariants, integrable systems and mirror symmetry. arXiv:1303.3253
[Mau1] D. Maulik, Stable pairs and the HOMFLY polynomial. arXiv:1210.6323
[MauOk] D. Maulik, A. Okounkov, Quantum groups and quantum cohomology. arXiv:1211.1287
[Nak1] H. Nakajima, AGT conjecture and perverse sheaves on instanton moduli spaces, in *Talk at the Simons Symposium on BPS States and Knot Invariants*, April (2012)
[Nak2] H. Nakajima, *Lectures on Hilbert Schemes of Points on Surfaces*. University Lecture Series (American Mathematical Society, Providence, RI, 1999)

[NagNak] K. Nagao, H. Nakajima, Counting invariant of perverse coherent sheaves and its wall-crossing. arXiv:0809.2992

[Neg] A. Negut, The shuffle algebra revisited. arXiv:1209.3349

[ORS] A. Oblomkov, J. Rasmussen, V. Shende, The Hilbert scheme of a plane singularity and the HOMFLY homology of its link. arXiv:1201.2115

[PaToVaVe] T. Pantev, B. Toen, M. Vaquié, G. Vezzosi, Shifted symplectic structures. arXiv:1111.3209

[PT1] R. Pandharipande, R.P. Thomas, Stable pairs and BPS invariants. arXiv:0711.3899

[Re1] M. Reineke, Framed quiver moduli, cohomology, and quantum groups. arXiv:math/0411101

[Re2] M. Reineke, The Harder-Narasimhan system in quantum groups and cohomology of quiver moduli. arXiv:math/0204059

[Rim] R. Rimanyi, On the Cohomological hall algebra of Dynkin quivers. arXiv:1303.3399

[Sch1] O. Schiffmann, Lectures on Hall algebras. arXiv:math/0611617

[SchV] O. Schiffmann, E. Vasserot, Cherednik algebras, W-algebras and equivariant cohomology of the moduli space of instantons on $\mathbb{A}^2$. arXiv: 1202.2756

[Se] P. Seidel, *Fukaya Categories and Picard-Lefschetz Theory*. Zurich Lectures in Advanced Mathematics (European Mathematical Society, Zurich, 2008)

[So1] Y. Soibelman, Talks at the Simons Symposium on BPS states and knot invariants, US Virgin Islands, April 2012 and AIM workshop on Donaldson-Thomas invariants and singularity theory, Budapest, May (2012)

[Sz1] B. Szendroi, Non-commutative Donaldson-Thomas theory and the conifold. arXiv:0705.3419

[Sz2] B. Szendroi, Nekrasov's partition function and refined Donaldson-Thomas theory: the rank one case. arXiv:1210.5181

[Tod] Y. Toda, Stability conditions and curve counting invariants on Calabi-Yau 3-folds. arXiv:1103.4229

[Xi] X. Xiao, The double of representations of Cohomological Hall Algebra for A_1-quiver. arXiv:1407.7593

A Stratification on the Moduli of K3 Surfaces in Positive Characteristic

Gerard van der Geer

Dedicated to the memory of Fritz Hirzebruch

Abstract We review the results on the cycle classes of the strata defined by the height and the Artin invariant on the moduli of K3 surfaces in positive characteristic obtained in joint work with Katsura and Ekedahl. In addition we prove a new irreducibility result for these strata.

1 Introduction

Moduli spaces in positive characteristic often possess stratifications for which we do not know characteristic 0 analogues. A good example is the moduli space of elliptic curves in characteristic $p > 0$. If E is an elliptic curve over an algebraically closed field k of characteristic p, then multiplication by p on E factors as

$$\times p : E \xrightarrow{F} E^{(p)} \xrightarrow{V} E,$$

where Frobenius F is inseparable and Verschiebung V can be separable or inseparable. If V is separable, then E is called ordinary, while if V is inseparable E is called supersingular. In the moduli space $\mathcal{A}_1 \otimes k$ of elliptic curves over k there are finitely many points corresponding to supersingular elliptic curves, and a well-known formula by Deuring, dating from 1941, gives their (weighted) number:

$$\sum_{E/k \text{ supers.}/\cong_k} \frac{1}{\#\mathrm{Aut}_k(E)} = \frac{p-1}{24},$$

where the sum is over the isomorphism classes of supersingular elliptic curves and each curve is counted with a weight. We thus find a stratification of the

G. van der Geer (✉)
Korteweg-de Vries Instituut, Universiteit van Amsterdam, Postbus 94248,
1090 GE Amsterdam, The Netherlands
e-mail: geer@science.uva.nl

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_14

moduli $\mathcal{A}_1 \otimes \mathbb{F}_p$ of elliptic curves with two strata: the ordinary stratum and the supersingular stratum. This stratification generalizes to the moduli of principally polarized abelian varieties of dimension g in positive characteristic where it leads to two stratifications, the Ekedahl-Oort stratification with 2^g strata and the Newton-polygon stratification. These stratifications have been the focus of much study in recent years (see, for example, [O, vdG, E-vdG1, Mo, R, C-O]). The dimension of these strata is known and in the case of the Ekedahl-Oort stratification we also know by [E-vdG1] the cycle classes of these strata in the Chow groups of a suitable compactification. The formulas for such cycle classes can be seen as a generalization of the formula of Deuring.

Besides abelian varieties, K3 surfaces form another generalization of elliptic curves. The stratification on the moduli of elliptic curves in positive characteristic generalizes to a stratification of the moduli $\mathcal{F}_g$ of primitively polarized K3 surfaces of degree $2g-2$ in positive characteristic. In fact, in the 1970s Artin and Mazur obtained in [A-M] an invariant of K3 surfaces by looking at the formal Brauer group of a K3 surface. For an elliptic curve the distinction between ordinary and supersingular can be formulated by looking at the formal group, that is, the infinite infinitesimal neighborhood of the origin with the inherited group law. If t is a local parameter at the origin, then multiplication by p is given by

$$[p]\, t = a\, t^{p^h} + \text{higher order terms} \tag{1}$$

with $a \neq 0$. Since multiplication by p on E is of degree p^2 and inseparable, we have $1 \leq h \leq 2$ and $h = 1$ if E is ordinary and $h = 2$ if E is supersingular. The formal group allows a functorial description as the functor on spectra S of local Artin k-algebras with residue field k given by

$$S \mapsto \ker\{H^1_{\text{et}}(E \times S, \mathbb{G}_m) \to H^1_{\text{et}}(E, \mathbb{G}_m)\},$$

where $H^1_{\text{et}}(E, \mathbb{G}_m) \cong H^1(E, \mathcal{O}_E^*)$ classifies line bundles on E. The invariant of Artin and Mazur generalizes this. For a K3 surface X they looked at the functor of local Artinian schemes over k with residue field k given by

$$S \mapsto \ker\{H^2_{\text{et}}(X \times S, \mathbb{G}_m) \to H^2_{\text{et}}(X, \mathbb{G}_m)\},$$

and showed that it is representable by a formal Lie group, called the formal Brauer group. Its tangent space is $H^2(X, \mathcal{O}_X)$, so we have a 1-dimensional formal group. Now over an algebraically closed field 1-dimensional formal groups are classified by their height: in terms of a local coordinate t multiplication by p is either zero or takes the form $[p]\, t = a\, t^{p^h} +$ higher order terms, with $a \neq 0$. If multiplication by p vanishes we say $h = \infty$, and then we have the formal additive group $\hat{\mathbb{G}}_a$ and if $h < \infty$ we have a p-divisible formal group.

Artin and Mazur connected this invariant $h(X)$ to the geometry of the K3 surface by proving that if $h(X) \neq \infty$ then

$$\rho(X) \leq 22 - 2h(X), \tag{2}$$

where $\rho(X)$ is the Picard number of X. In particular, either we have $\rho = 22$ (and then necessarily $h = \infty$), or $\rho \leq 20$.

The case that $\rho = 22$ can occur in positive characteristic as Tate observed: for example, in characteristic $p \equiv 3 \pmod 4$ the Fermat surface $x^4 + y^4 + z^4 + w^4 = 0$ has $\rho = 22$ (see [T, S]).

If $h(X) = \infty$, then the K3 surface X is called supersingular. By the result of Artin and Mazur a K3 surface X with $\rho(X) = 22$ must be supersingular. In 1974 Artin conjectured the converse ([A]): a supersingular K3 surface has $\rho(X) = 22$. This has now been proved by Maulik, Charles, and Madapusi Pera for $p \geq 3$, see [M, C, P2]. In the 1980s Rudakov, Shafarevich, and Zink proved that supersingular K3 surfaces with a polarization of degree 2 have $\rho = 22$ in characteristic $p \geq 5$, see [R-S-Z].

The height is upper semi-continuous in families. The case $h = 1$ is the generic case; in particular, the K3 surfaces with $h = 1$ form an open set. By the inequality (2) we have

$$1 \leq h \leq 10 \qquad \text{or} \quad h = \infty.$$

In the moduli space $\mathcal{F}_g$ of primitively polarized K3 surfaces of genus g (or equivalently, of degree $2g-2$) with $2g-2$ prime to p, the locus of K3 surfaces with height $\geq h$ is locally closed and has codimension $h-1$ and we thus have 11 strata in the 19-dimensional moduli space $\mathcal{F}_g$. The supersingular locus has dimension 9. Artin showed that it is further stratified by the Artin invariant σ_0: assuming that $\rho = 22$ one looks at the Néron–Severi group $\mathrm{NS}(X)$ with its intersection pairing; it turns out that the discriminant group $\mathrm{NS}(X)^\vee/\mathrm{NS(X)}$ is an elementary p-group isomorphic to $(\mathbb{Z}/p\mathbb{Z})^{2\sigma_0}$ and one thus obtains another invariant. The idea behind this is that, though $\rho = 22$ stays fixed, divisor classes in the limit might become divisible by p, thus changing the Néron–Severi lattice and σ_0. The invariant σ_0 is lower semi-continuous. The generic case (supersingular) is where $\sigma_0 = 10$ and the most degenerate case is the so-called superspecial case $\sigma_0 = 1$.

In total one obtains a stratification on the moduli space $\mathcal{F}_g$ of K3 surfaces with a primitive polarization of genus g with 20 strata V_j with $1 \leq j \leq 20$

$$\overline{V}_j = \{[X] \in \mathcal{F}_g : h(X) \geq j\} \qquad 1 \leq j \leq 10,$$

and

$$\overline{V}_j = \{[X] \in \mathcal{F}_g : h(X) = \infty,\, \sigma_0(X) \leq 21 - j\} \qquad 11 \leq j \leq 20\,,$$

the closures $\overline{V}_j$ of which are linearly ordered by inclusion. In joint work with Katsura [vdG-K1] we determined the cycle classes of the strata $\overline{V}_j$ ($j = 1, \ldots, 10$) of height $h \geq j$

$$[V_j] = (p-1)(p^2-1)\cdots(p^{j-1}-1)\,\lambda_1^{j-1}, \qquad 1 \le j \le 10 \tag{3}$$

where $\lambda_1 = c_1(\pi_*(\Omega^2_{\mathcal{X}/\mathcal{F}_g}))$ is the Hodge class with $\pi : \mathcal{X} \to \mathcal{F}_g$ the universal family. We also determined the class of the supersingular locus $\overline{V}_{11}$. Moreover, we proved that the singular locus of $\overline{V}_j$ is contained in the stratum of the supersingular locus where the Artin invariant is at most $j-1$, see [vdG-K1, Theorem 14.2]. Ogus made this more precise in [Og2]. For more on the moduli of supersingular K3 surfaces we also refer to [Og1, R-S2, L].

But the cycle classes of the other strata $\overline{V}_j$ for $j = 12, \ldots, 20$ [A-SD, H, L, Og1, R-S, R-S2] given by the Artin invariant turned out to be elusive. In joint work with Ekedahl [E-vdG2] we developed a uniform approach by applying the philosophy of [vdG, E-vdG1] of interpreting these stratifications in terms of flags on the cohomology and eventually were able to determine all cycle classes. All these cycle classes are multiples of powers of the Hodge class λ_1.

Our approach uses flags on the de Rham cohomology, here on H^2_{dR} as opposed to H^1_{dR} for abelian varieties. The space $H^2_{\mathrm{dR}}(X)$ is provided with a non-degenerate intersection form and it carries a filtration, the Hodge filtration. But in positive characteristic it carries a second filtration deriving from the fact that we do not have a Poincaré lemma, or in other words, it derives from the Leray spectral sequence applied to the relative Frobenius morphism $X \to X^{(p)}$. See later for more on this so-called conjugate filtration. We thus find two filtrations on $H^2_{\mathrm{dR}}(X)$ and these are not necessarily transversal. We say that X is *ordinary* if the two filtrations are transversal and that X is *superspecial* if the two filtrations coincide. These are two extremal cases, but by considering the relative position of flags refining the two flags one obtains a further discrete invariant and one retrieves in a uniform way the invariants encountered above, the height h and the Artin invariant σ_0.

For applications it is important that we consider moduli of K3 surfaces together with an embedding of a non-degenerate lattice N in the Néron–Severi group of X such that it contains a semi-ample class of degree prime to the characteristic p, and then look at the primitive part of the de Rham cohomology. If the dimension n of this primitive cohomology is even, this forces us to deal with very subtle questions related to the distinction of orthogonal group $\mathrm{O}(n)$ versus the special orthogonal group $\mathrm{SO}(n)$.

Instead of working directly on the moduli spaces of K3 surfaces, we work on the space of flags on the primitive part of H^2_{dR}, that is, we work on a flag bundle over the moduli space. The reason is that the strata that are defined on this space are much better behaved than the strata on the moduli of K3 surfaces itself. In fact, up to infinitesimal order p the strata on the flag space over $\mathcal{F}_g$ look like the strata (the Schubert cycles) on the flag space for the orthogonal group. These strata are indexed by elements of a Weyl group.

In order to get the cycle classes of the strata on the moduli of K3 surfaces we note that these latter strata are linearly ordered. This allows us to apply fruitfully a Pieri type formula which expresses the intersection product of a cycle class with a first Chern class (the Hodge class in our case) as a sum of cycle classes of one dimension less.

We apply this on the flag space and then project it down. There are many more strata on the flag space of the primitive cohomology than on the moduli space. Some of these strata, the so-called final ones, map in an étale way to their image in the moduli space; for the non-final ones, either the image is lower-dimensional, and hence its cycle class can be ignored, or the map is inseparable and factors through a final stratum and the degree of the inseparable map can be calculated. In this way one arrives at closed formulas for the cycle classes of the strata on the moduli space.

We give an example of the formula for the cycle classes from [E-vdG2] in the following special case. Let $p > 2$ and $\pi : \mathcal{X} \to \mathcal{F}_g$ be the universal family of K3 surfaces with a primitive polarization of degree $d = 2g - 2$ with d prime to p. Then there are 20 strata on the 19-dimensional moduli space $\mathcal{F}_g$ parametrized by so-called final elements w_i with $1 \leq i \leq 20$ in the Weyl group of SO(21). These are ordered by their length $\ell(w_i)$ (in the sense of length in Weyl groups) starting with the longest one. The strata $\mathcal{V}_{w_i}$ for $i = 1, \dots 10$ are the strata of height $h = i$, the stratum $\mathcal{V}_{w_{11}}$ is the supersingular stratum, while the strata $\mathcal{V}_{w_i}$ for $i = 11, \dots, 20$ are the strata where the Artin invariant satisfies $\sigma_0 = 21 - i$.

Theorem 1.1. *The cycle class of the closed stratum $\overline{\mathcal{V}}_{w_i}$ on the moduli space $\mathcal{F}_g$ is given by*

$$\text{(i)}\quad [\overline{\mathcal{V}}_{w_k}] = (p-1)(p^2-1)\cdots(p^{k-1}-1)\lambda_1^{k-1} \quad \textit{if } 1 \leq k \leq 10,$$

$$\text{(ii)}\quad [\overline{\mathcal{V}}_{w_{11}}] = \frac{1}{2}(p-1)(p^2-1)\cdots(p^{10}-1)\lambda_1^{10},$$

$$\text{(iii)}\quad [\overline{\mathcal{V}}_{w_{10+k}}] = \frac{1}{2}\frac{(p^{2k}-1)(p^{2(k+1)}-1)\cdots(p^{20}-1)}{(p+1)\cdots(p^{11-k}+1)}\lambda_1^{9+k} \quad \textit{if } 2 \leq k \leq 10.$$

Here $\lambda_1 = c_1(L)$ with $L = \pi_*(\Omega^2_{\mathcal{X}/\mathcal{F}_g})$ is the Hodge class. Sections of $L^{\otimes r}$ correspond to modular forms of weight r. It is known (cf. [vdG-K2]) that the class $\lambda_1^{18} \in \mathrm{CH}^{18}_{\mathbb{Q}}(\mathcal{F}_g)$ vanishes on $\mathcal{F}_g$. But the formulas can be made to work also on the closure of the image $\mathcal{F}_g$ embedded in projective space by the sections of a sufficiently high power of L, so that the last two formulas (involving λ_1^{18} and λ_1^{19}) are non-trivial and still make sense. Note here that λ_1 is an ample class; this is well known by Baily-Borel in characteristic 0, but now we know it too in characteristic $p \geq 3$ by work of Madapusi-Pera [P1] and Maulik [M].

In particular, we can give an explicit formula for the weighted number of superspecial K3 surfaces of genus g by using a formula for $\deg(\lambda_1^{19})$ from [G-H-S]. We consider the situation where we have a primitive polarization of degree $d = 2d'$ inside N with

$$N^{\perp} = 2\,U \perp m\,E_8(-1) \perp \langle -d \rangle\,, \tag{4}$$

where U is a hyperbolic plane and $m = 0$ or $m = 2$.

Theorem 1.2. *The weighted number*

$$\sum_{X \text{ superspecial}/\cong} \frac{1}{\#\mathrm{Aut}_k(X)}$$

of superspecial K3 surfaces with a primitive polarization N of degree $d = 2d'$ prime to the characteristic p with $N^{\perp}$ as in (1), is given by

$$\frac{-1}{2^{4m+1}} \frac{p^{8m+4}-1}{p+1} \left((d')^{10} \prod_{\ell \mid d'} (1+\ell^{-4m-2}) \right) \zeta(-1)\zeta(-3)\cdots\zeta(-8m-3)$$

where ζ denotes the Riemann zeta function and ℓ runs over the primes dividing d'.

For $m = 0$ this formula can be applied to count the number of Kummer surfaces coming from superspecial principally polarized abelian surfaces and the formula then agrees with the formulas of [E, vdG].

Formulas like those given in (3) and in Theorem 1.1 for the classes of the height strata were obtained in joint work with Katsura [vdG-K1] by different (ad hoc) methods using formal groups and Witt vector cohomology; but these methods did not suffice to calculate the cycle classes of the Artin invariant strata.

A simple corollary is (see [E-vdG2, Proposition 13.1]).

Corollary 1.3. *A supersingular (quasi-)elliptic K3 surface with a section cannot have Artin invariant $\sigma_0 = 10$.*

This result was obtained independently by Kondo and Shimada using a different method in [K-S, Corollary 1.6].

In addition to reviewing the results from [E-vdG2] we prove irreducibility results for the strata; about half of the strata on the moduli space $\mathcal{F}_g$ are shown to be irreducible. Here we use the local structure of the strata on the flag space.

Theorem 1.4. *Let $p \geq 3$ prime to the degree $d = 2g - 2$. For a final element $w \in W_m^B$ (resp. $w \in W_m^D$) of length $\ell(w) \geq m$ the stratum $\overline{\mathcal{V}}_w$ in $\mathcal{F}_g$ is irreducible.*

So the strata above the supersingular locus are all irreducible. We have a similar result in $\mathcal{F}_N$.

The formulas we derived deal with the group $\mathrm{SO}(n)$; for K3 surfaces we can restrict $n \leq 21$, but the formulas for larger n might find applications to the moduli of hyperkähler varieties in positive characteristic (by looking at the middle dimensional de Rham cohomology or at H^2_{dR} equipped with the Beauville-Bogomolov form).

2 Filtrations on the de Rham Cohomology of a K3 Surface

Let X be a K3 surface over an algebraically closed field of characteristic $p > 2$ and let $N \hookrightarrow \mathrm{NS}(X)$ be an isometric embedding of a non-degenerate lattice in the Néron–Severi group $\mathrm{NS}(X)$ (equal to the Picard group for a K3 surface) and assume that N contains a semi-ample line bundle and that the discriminant of N is coprime with p (that is, p does not divide $\#N^\vee/N$). We let $N^\perp$ be the primitive cohomology, that is, the orthogonal complement of the image of $c_1(N)$ of N in $H^2_{\mathrm{dR}}(X)$. It carries a Hodge filtration

$$0 = U_{-1} \subset U_0 \subset U_1 \subset U_2 = N^\perp$$

of dimensions $0, 1, n-1, n$ and comes with a non-degenerate intersection form for which the Hodge filtration is self-dual: $U_0^\perp = U_1$. Now in positive characteristic we have another filtration

$$0 = U^c_{-1} \subset U^c_0 \subset U^c_1 \subset U^c_2 = N^\perp\,,$$

the conjugate filtration; it is self-dual too. The reason for its existence is that the Poincaré lemma does not hold in positive characteristic. If $F : X \to X^{(p)}$ is the relative Frobenius morphism, then we have a canonical (Cartier) isomorphism

$$C : \mathcal{H}^j(F_*\Omega^\bullet_{X/k}) \cong \Omega^j_{X^{(p)}/k}$$

and we get a non-trivial spectral sequence from this: the second spectral sequence of hypercohomology with E_2-term $E_2^{ij} = H^i(X^{(p)}, \mathcal{H}^j(\Omega^\bullet))$ which by the inverse Cartier isomorphism $C^{-1} : \Omega^j_{X^{(p)}} \simeq \mathcal{H}^j(F_*(\Omega^\bullet_{X/k}))$ can be rewritten as $H^i(X^{(p)}, \Omega^j_{X^{(p)}/k})$, degenerating at the E_2-term and abutting to $H^{i+j}_{\mathrm{dR}}(X/k)$. This leads to a second filtration on the de Rham cohomology.

The inverse Cartier operator gives an isomorphism

$$F^*(U_i/U_{i-1}) \cong U^c_{2-i}/U^c_{1-i}\,.$$

As a result we have two (incomplete) flags forming a so-called F-zip in the sense of [M-W]. Unlike the characteristic zero situation where the Hodge flag and its complex conjugate are transversal, the two flags in our situation are not necessarily transversal. In fact, the K3 surface X is called ordinary if these flags are transversal and superspecial if they coincide. These are just two extremal cases among more possibilities.

Before we deal with these further possibilities, we recall some facts about isotropic flags in a non-degenerate orthogonal space. Let V be a non-degenerate orthogonal space of dimension n over a field of characteristic $p > 2$. We have to distinguish the cases n odd and n even, the latter being more subtle. We look at isotropic flags

$$(0) = V_0 \subset V_1 \subset \cdots \subset V_r$$

with $\dim V_i = i$ in V, that is, we require that the intersection form vanishes on V_r. We call the flag maximal if $r = [n/2]$. We can complete a maximal flag by putting $V_{n-j} = V_j^{\perp}$. Now if $n = 2m$ is even, a complete isotropic flag $V_\bullet$ determines another complete isotropic flag by putting $V_i' = V_i$ for $i < n/2$ and by taking for V_m' the unique maximal isotropic space containing V_{m-1} but different from V_m. We call this flag $V_\bullet'$ the *twist* of $V_\bullet$.

In fact, if n is even, the group $\mathrm{SO}(n)$ does not act transitively on complete flags.

Given two complete isotropic flags their relative position is given by an element of a Weyl group. If n is odd, we let W_m^B be the Weyl group of $\mathrm{SO}(2m+1)$. It can be identified with the following subgroup of the symmetric group $\mathfrak{S}_{2m+1}$

$$\{\sigma \in \mathfrak{S}_{2m+1} : \sigma(i) + \sigma(2m+2-i) = 2m+2 \text{ for all } 1 \le i \le 2m+1\}.$$

The fact is now that the $\mathrm{SO}(2m+1)$-orbits of pairs of totally isotropic complete flags are in 1-1 correspondence with the elements of W_m^B given by

$$w \longleftrightarrow \Big(\sum_{j \le i} k \cdot e_j, \sum_{j \le i} k \cdot e_{w^{-1}(j)}\Big)$$

with the e_i a fixed orthogonal basis with $\langle e_i, e_j \rangle = \delta_{i,2m+2-j}$. The simple reflections $s_i \in W_m^B$ for $i = 1, \ldots, m$ are given by $s_i = (i, i+1)(2m+1-i, 2m+2-i)$ if $i < m$ and by $s_m = (m, m+2)$, and will play an important role here.

But in the case that $n = 2m$ is even we have to replace the Weyl group W_m^C (of $\mathrm{O}(2m)$) given by

$$\big\{\sigma \in \mathfrak{S}_{2m} : \sigma(i) + \sigma(2m+1-i) = 2m+1 \text{ for all } 1 \le i \le 2m\big\}$$

by the index 2 subgroup W_m^D given by the extra parity condition

$$\#\{1 \le i \le m : \sigma(i) > m\} \equiv 0 \,(\mathrm{mod} 2).$$

The simple reflections $s_i \in W_m^D$ are given by $s_i = (i, i+1)(2m-i, 2m+1-i)$ for $i < m$ and by $s_m = (m-1, m+1)(m, m+2)$. In the larger group W_m^C we have the simple reflections s_i with $1 \le i \le m-1$ and $s_m' = (m, m+1)$. Note that s_m' commutes with the s_i for $i = 1, \ldots, m-2$ and conjugation by it interchanges s_{m-1} and s_m.

The $\mathrm{SO}(2m)$-orbits of pairs of totally isotropic complete flags are in bijection with the elements of W_m^C given by

$$w \longleftrightarrow \Big(\sum_{j \le i} k \cdot e_j, \sum_{j \le i} k \cdot e_{w^{-1}(j)}\Big)$$

with basis e_i with $\langle e_i, e_j \rangle = \delta_{i,2m+1-j}$. Twisting the first (resp. second) flag corresponds to changing w to ws_m' (resp. $s_m' w$).

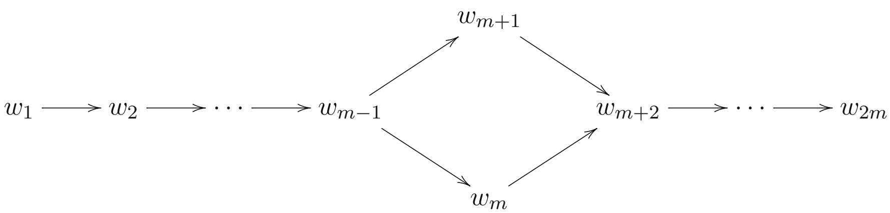
Back to K3 surfaces. We can refine the conjugate flag on $H^2_{\mathrm{dR}}(X)$ to a full (increasing) flag $D^\bullet$ and use the Cartier operator to transfer it back to a decreasing flag $C^\bullet$ on the Hodge filtration $U_\bullet$. We thus get two full flags.

Definition 2.1. A full flag refining the conjugate filtration is called stable if $D_j \cap U_i + U_{i-1}$ is an element of the $C^\bullet$ filtration or of its twist. A flag is called *final* if it is stable and complete.

Final flags correspond to the so-called *final elements* in the Weyl group defined as follows. Elements in the Weyl group W_m^B (resp. W_m^D) which are reduced with respect to the set of roots obtained after removing the first root [so that the remaining roots form a root system of type B_{m-1} (resp. D_{m-1})] are called final elements.

If $n = 2m + 1$ is odd, we have $2m$ final elements in W_m^B. These are the elements σ given by $[\sigma(1), \sigma(2), \ldots, \sigma(m)]$ and we can list these as $w_1 = [2m + 1, 2, 3, \ldots]$, $w_2 = [2m, 1, 3, \ldots], \ldots, w_{2m} = [1, 2, \ldots, m]$. These final elements w_j are linearly ordered by their length $\ell(w_j) = 2m - j$, with w_1 being the longest element and w_{2m} equal to the identity element.

If $n = 2m$, then we also have $2m$ final elements in W_m^C, but these are no longer linearly ordered by their length, but the picture is rather



Conjugation by s'_m interchanges the two final elements of length $m - 1$. This corresponds to twisting a complete isotropic flag. We shall denote $w_j s'_m$ by w'_j and we shall call these *twisted final elements.*

The following theorem shows that we can read off the height $h(X)$ of the formal Brauer group and the Artin invariant $\sigma_0(X)$ from these final filtrations on H^2_{dR}. The following theorem is proven in [E-vdG2]. Recall that the discriminant of N is assumed to be prime to p.

Theorem 2.2. *Let X be a K3 surface with an embedding $N \hookrightarrow \mathrm{NS}(X)$ and let $H \subset H^2_{\mathrm{dR}}$ be the primitive part of the cohomology with $n = \dim(H)$ and $m = [n/2]$. Then H possesses a final filtration; all final filtrations are of the same combinatorial type w. Moreover,*

(i) *X has finite height $h < n/2$ if and only if $w = w_h$ or w'_h.*
(ii) *X has finite height $h = n/2$ if and only if $w = w'_m$.*
(iii) *X has Artin invariant $\sigma_0 < n/2$ if and only if $w = w_{2m+1-\sigma_0}$ or $w = w'_{2m+1-\sigma_0}$.*
(iv) *X has Artin invariant $\sigma_0 = n/2$ if and only if $w = w_{m+1}$.*

In case (i) [resp. in case (iii)] we can distinguish these cases $w = w_h$ or $w = w_h'$ (resp. $w = w_{n-\sigma_0}$ or $w = w'_{n-\sigma_0}$) for even n by looking whether the so-called middle part of the cohomology is split, or equivalently, by the sign of the permutation w. We get w_h' in case (i) exactly if w is an odd permutation and in case (iii) we get $w = w_{n-\sigma_0}$ exactly if w is an even permutation.

Looking at the diagram above one sees that the theorem excludes one of the two possibilities corresponding to the two final elements w_m and w_{m+1} of length m. This is analyzed in detail in [E-vdG2, Sect. 5]. It then agrees with the fact that the (closed) strata defined by the height and the Artin invariant are linearly ordered by inclusion, whereas the final w_i in the above diagram are not.

3 Strata on the Flag Space

Suppose that we have a family $f : \mathcal{X} \to S$ of N-marked K3 surfaces over a smooth base S. We shall make a versality assumption. At a geometric point s of S we have the Kodaira-Spencer map $T_sS \to H^1(X_s, T^1_{X_s})$. We have a natural map $H^1(X_s, T^1_{X_s}) \to \mathrm{Hom}(H^0(X_s, \Omega^2_{X_s}), H^1(X_s, \Omega^1_{X_s}))$ and we can project $H^1(X_s, \Omega^1_{X_s})$ to the orthogonal complement P of the image of $N \hookrightarrow \mathrm{NS}(X_s)$ in $H^1(X_s, \Omega^1_{X_s})$.

Assumption 3.1. The versality assumption is the requirement that the resulting map

$$T_sS \to \mathrm{Hom}(H^0(X_s, \Omega^2_{X_s}), P)$$

is surjective.

The primitive cohomology forms a vector bundle $\mathcal{H}$ of rank n over S. It comes with two partial orthogonal isotropic flags: the conjugate flag and the Hodge flag. If we choose a complete orthogonal flag refining the conjugate filtration and transfer it to the Hodge filtration by the Cartier operator we get two flags and we can measure the relative position. This defines strata on S. This implies that we have to choose a flag and we are thus forced to work on the flag space $\mathcal{B}$ over S (or $\mathcal{B}_N$ over $\mathcal{F}_N$) of complete isotropic flags refining the Hodge filtration. (Since we are using $\mathcal{F}_N$ for the moduli space of N-polarized K3 surfaces we use another letter for the flag space; say $\mathcal{B}_N$ for banner).

To define the strata scheme-theoretically we consider the general case of a semi-simple Lie group G and a Borel subgroup B and a G/B-bundle $R \to Y$ over some scheme Y with G as structure group. Let $r_i : Y \to R$ $(i = 1, 2)$ be two sections. If w is an element of the Weyl group W of G we define a locally closed subscheme $\mathcal{U}_w$ of Y as follows. We choose locally (possibly in the étale topology) a trivialization of R such that r_1 is a constant section. Then r_2 corresponds to a map $Y \to G/B$ and we define $\mathcal{U}_w$ (resp. $\overline{\mathcal{U}}_w$) to be the inverse image of the B-orbit BwB (resp. of its closure).

We thus find strata $\mathcal{U}_w$ and $\overline{\mathcal{U}}_w$ of $\mathcal{B}_N$; it turns out that $\overline{\mathcal{U}}_w$ is the closure of $\mathcal{U}_w$. It might seem that working on the flag space brings us farther from the goal of defining and studying strata on the base space S or on the moduli spaces. However, working with the strata on the flag space has the advantage that the strata are much better behaved there.

The space $\mathcal{B}_N$ together with the strata $\mathcal{U}_w$ is a stratified space. The space $\mathcal{F}l_n$ of complete self-dual flags on an orthogonal space V also carries a stratification, namely by Schubert cells. It is fiber space over the space of maximal isotropic subspaces $\mathcal{I}_n$.

The main idea is now that our space $\mathcal{B}_N$ over $\mathcal{F}_N$ locally at a point up to the $(p-1)$st infinitesimal neighborhood looks like $\mathcal{F}l_n \to \mathcal{I}_n$ at a suitable point as stratified spaces. This idea was developed in [E-vdG1] and here it profitably can be used too.

If (R, m) is a local ring *the height* 1 *hull* of R (resp. of $S = \mathrm{Spec}(R)$) is $R/m^{(p)}$ (resp. $\mathrm{Spec}(R/m^{(p)})$) with $m^{(p)}$ the ideal generated by the pth powers of elements of m. It defines the height 1 neighborhood of the point given by m. We call two local rings height 1-isomorphic if their height 1 hulls are isomorphic.

Theorem 3.2. *Let k be a perfect field of characteristic p. For each k-point x of $\mathcal{B}_N$ there exists a k-point y of $\mathcal{F}l_n$ such that the height* 1 *neighborhood of x is isomorphic (as stratified spaces) to the height* 1 *neighborhood of y.*

Indeed, we can trivialize the de Rham cohomology with its Gauss–Manin connection on the height 1 neighborhood of x (because the ideal of x has a divided power structure for which divided powers of degree $\geq p$ are zero). This has strong consequences for our strata, cf. the following result from [E-vdG2].

Theorem 3.3. *The strata $\mathcal{U}_w$ on the flag space $\mathcal{B}_N$ satisfy the following properties:*

(1) *The stratum $\mathcal{U}_w$ is smooth of dimension equal to the length $\ell(w)$ of w.*
(2) *The closed stratum $\overline{\mathcal{U}}_w$ is reduced, Cohen-Macaulay and normal of dimension $\ell(w)$ and equals the closure of $\mathcal{U}_w$.*
(3) *If w is a final element, then the restriction of $\mathcal{B}_N \to \mathcal{F}_N$ to $\mathcal{U}_w$ is a finite surjective étale covering from $\mathcal{U}_w$ to its image $\mathcal{V}_w$.*

The degrees of the maps $\pi_w : \mathcal{U}_w \to \mathcal{V}_w$ for final w coincide with the number of final filtrations of type w and these numbers can be calculated explicitly. For example, for $w_i \in W_m^B$ with $1 \leq i < m$ we have

$$\deg \pi_{w_i} / \deg \pi_{w_{i+1}} = p^{2m-2i-1} + p^{2m-2i-2} + \cdots + 1.$$

4 The Cycle Classes

We consider a family of N-polarized K3 surfaces $\mathcal{X} \to S$ with S smooth and satisfying the versality Assumption 3.1. Our strategy in [E-vdG2] is to apply inductively a Pieri formula to the final strata, which expresses the intersection

$\lambda_1 \cdot [\overline{\mathcal{U}}_w]$ as a sum over the classes $[\overline{\mathcal{U}}_v]$, where v is running through the elements of the Weyl group of the form $v = ws_\alpha$ with s_α simple and $\ell(ws_\alpha) = \ell(w) - 1$. In fact, we use a Pieri formula due to Pittie and Ram [P-R]. In general these elements v of colength 1 are not final and this forces us to analyze what happens with the strata $\mathcal{U}_v$ under the projection $\mathcal{B}_N \to \mathcal{F}_N$. It turns out that for a non-final stratum either the projection is to a lower-dimensional stratum or factors through an inseparable map to a final stratum. The degree of these inseparable maps can be calculated. In the case of a map to a lower dimensional stratum we can neglect these for the cycle class calculation.

So suppose that for an element w in the Weyl group we have $\ell(ws_i) = \ell(w) - 1$ for some $1 < i \leq m$. This means that if $A_\bullet$ and $B_\bullet$ denote the two flags, that the image of $B_{w(i+1)} \cap A_{i+1}$ in A_{i+1}/A_i is 1-dimensional and thus we can change the flag $A_\bullet$ to a flag $A'_\bullet$ by setting $A'_j = A_j$ for $j \neq i$ and A'_i/A_{i-1} equal to the image of $B_{w(i+1)} \cap A_{i+1}$. This gives us a map

$$\sigma_{w,i} : \mathcal{U}_w \to \mathcal{F}_N, \qquad (A_\bullet, B_\bullet) \mapsto (A'_\bullet, B_\bullet).$$

In this situation the image depends on the length $\ell(s_i w s_i)$:

Lemma 4.1. *If $\ell(s_i w s_i) = \ell(w)$, then the image of $\sigma_{w,i}$ is equal to $\mathcal{U}_{s_i w s_i}$ and the map is purely inseparable of degree p. If $\ell(s_i w s_i) = \ell(w) - 2$, then $\sigma_{w,i}$ maps onto $\mathcal{U}_{ws_i} \cup \mathcal{U}_{s_i w s_i}$ and the map $\sigma_{w,i}$ is not generically finite.*

We then analyze in detail the colength 1 elements occurring and whether they give rise to projections that lose dimension or are inseparable to final strata. This is carried out in detail in [E-vdG2, Sects. 9–11]. In this way the Pieri formula enables us to calculate the cycle classes.

The result for the cycle classes of the strata $\overline{\mathcal{V}}_{w_i}$ in the case that n is odd (and with $m = [n/2]$) reads (cf. [E-vdG2]):

Theorem 4.2. *The cycle classes of the final strata $\overline{\mathcal{V}}_w$ on the base S are polynomials in λ_1 with coefficients that are polynomials in $\frac{1}{2}\mathbb{Z}[p]$ given by*

$$\text{(i)} \quad [\overline{\mathcal{V}}_{w_k}] = (p-1)(p^2-1)\cdots(p^{k-1}-1)\lambda_1^{k-1} \quad \textit{if } 1 \leq k \leq m,$$

$$\text{(ii)} \quad [\overline{\mathcal{V}}_{w_{m+1}}] = \frac{1}{2}(p-1)(p^2-1)\cdots(p^m-1)\lambda_1^m,$$

$$\text{(iii)} \quad [\overline{\mathcal{V}}_{w_{m+k}}] = \frac{1}{2}\frac{(p^{2k}-1)(p^{2(k+1)}-1)\cdots(p^{2m}-1)}{(p+1)\cdots(p^{m-k+1}+1)}\lambda_1^{m+k-1} \quad \textit{if } 2 \leq k \leq m.$$

In the case where n is even the result for the untwisted final elements is the following:

Theorem 4.3. *The cycle classes of the final strata $\overline{\mathcal{V}}_w$ for final elements $w = w_j \in W_m^D$ on the base S are powers of λ_1 times polynomials in $\frac{1}{2}\mathbb{Z}[p]$ given by*

$$\text{(i)}\quad [\overline{\mathcal{V}}_{w_k}] = (p-1)(p^2-1)\cdots(p^{k-1}-1)\lambda_1^{k-1} \quad \text{if } k \le m-1,$$

$$\text{(ii)}\quad [\overline{\mathcal{V}}_{w_{m+1}}] = (p-1)(p^2-1)\cdots(p^{m-1}-1)\lambda_1^{m-1},$$

$$\text{(iii)}\quad [\overline{\mathcal{V}}_{w_{m+k}}] = \frac{1}{2}\frac{\prod_{i=1}^{m-1}(p^i-1)\prod_{i=m-k+2}^{m}(p^i+1)}{\prod_{i=1}^{k-2}(p^i+1)\prod_{i=1}^{k-1}(p^i-1)}\lambda_1^{m+k-2} \quad \text{if } 2 \le k \le m.$$

Furthermore, we have that $\overline{\mathcal{V}}_{w_m} = \emptyset$.

Finally in the twisted even case we have

Theorem 4.4. *The cycle classes of the final strata* $\overline{\mathcal{V}}_w$ *for twisted final elements* $w = w_j \in W_m^D s_m'$ *on the base* S *are powers in* λ_1 *with coefficients that are polynomials in* $\frac{1}{2}\mathbb{Z}[p]$ *given by*

$$\text{(i)}\quad [\overline{\mathcal{V}}_{w_k}] = (p-1)(p^2-1)\cdots(p^{k-1}-1)\lambda_1^{k-1} \quad \text{if } k \le m-1,$$

$$\text{(ii)}\quad [\overline{\mathcal{V}}_{w_m}] = (p-1)(p^2-1)\cdots(p^{m}-1)\lambda_1^{m-1},$$

$$\text{(iii)}\quad [\overline{\mathcal{V}}_{w_{m+k}}] = \frac{1}{2}\frac{\prod_{i=1}^{m}(p^i-1)\prod_{i=m-k+2}^{m-1}(p^i+1)}{\prod_{i=1}^{k-1}(p^i+1)\prod_{i=1}^{k-2}(p^i-1)}\lambda_1^{m+k-2} \quad \text{if } 2 \le k \le m.$$

Furthermore, we have $\overline{\mathcal{V}}_{w_{m+1}} = \emptyset$.

5 Irreducibility

In this section we shall show that about half of the $2m$ strata $\mathcal{V}_{w_i}$ on our moduli space $\mathcal{F}_N$ of N-polarized K3 surfaces are irreducible (m strata in the B-case, $m-1$ in the D-case).

Theorem 5.1. *Let* $p \ge 3$ *and assume that* $\mathcal{F}_N$ *is the moduli space of primitively* N*-polarized K3 surfaces where* $N^\vee/N$ *has order prime to* p*. If* $w \in W_m^B$ *(resp.* $w \in W_m^D$ *or* $w \in W_m^D s_m'$*) is a (twisted) final element with length* $\ell(w) \ge m$*, then the locus* $\overline{\mathcal{V}}_w$ *in* $\mathcal{F}_N$ *is irreducible.*

Proof. (We do the B-case, leaving the other case to the reader.) The idea behind the proof is to show that for $1 \le i \le m$ the stratum $\overline{\mathcal{U}}_{w_i}$ is connected in the flag space $\mathcal{B}_N$. Note that $\mathcal{F}_N$ is connected by our assumptions. By Theorem 3.3 the stratum $\overline{\mathcal{U}}_{w_i}$ is normal, so if it is connected it must be irreducible. But then its image $\overline{\mathcal{V}}_{w_i}$ in $\mathcal{F}_N$ is irreducible as well. This shows the advantage of working on the flag space.

To show that $\overline{\mathcal{U}}_{w_i}$ is connected in $\mathcal{B}_N$ we use that its 1-skeleton is connected, that is, that the union of the 1-dimensional strata that it contains, is connected and that every irreducible component of $\overline{\mathcal{U}}_{w_i}$ intersects the 1-skeleton. To do that we prove the following facts:

(1) The loci $\overline{\mathcal{V}}_{w_i}$ in $\mathcal{F}_N$ are connected for $i < 2m$ (that is, for $w_i \ne 1$).
(2) Any irreducible component of any $\overline{\mathcal{U}}_w$ contains a point of $\mathcal{U}_1$.

(3) The union $\cup_{i=2}^m \overline{\mathcal{U}}_{s_i}$ intersected with a fiber of $\mathcal{B}_N \to \mathcal{F}_N$ over a point of the superspecial locus $\mathcal{V}_1$ is connected.
(4) For $1 \leq i \leq m$ the locus $\overline{\mathcal{U}}_{w_i}$ contains $\cup_{i=1}^m \overline{\mathcal{U}}_{s_i}$.

In the proof we use the fact that the closure of strata on the flag space is given by the Bruhat order in the Weyl group: $\mathcal{U}_v$ occurs in the closure of $\mathcal{U}_w$ if $v \geq w$ in the Bruhat order. Furthermore, we observe that one knows by [M, P1] that λ_1 is an ample class.

Together (1) and (3) will prove that the locus $\cup_{i=1}^m \overline{\mathcal{U}}_{s_i}$ (whose image in $\mathcal{F}_N$ is $\overline{\mathcal{V}}_{2m-1}$) is connected. We begin by proving (1).

Sections of a sufficiently high multiple of λ_1 embed $\mathcal{F}_N$ into projective space and we take its closure $\overline{\mathcal{F}}_N$. By the result of Theorem 4.2 (resp. 4.3 and 4.4) we know that the cycle class $[\overline{\mathcal{V}}_{w_i}]$ is a multiple of λ_1^{i-1}, so these loci are connected in $\overline{\mathcal{F}}_N$ for $i-1 < \dim \mathcal{F}_N$. In particular, the 1-dimensional locus $\overline{\mathcal{V}}_{w_{2m-1}}$ in $\mathcal{F}_N$ (which equals its closure in $\overline{\mathcal{F}}_N$) is connected. On any irreducible component of Y of $\overline{\mathcal{V}}_{w_i}$ in $\overline{\mathcal{F}}_N$ with $i < 2m-1$ the intersection with $\overline{\mathcal{V}}_{w_{2m-1}}$ is cut out by a multiple of a positive power of λ_1, hence it intersects this locus (in $\mathcal{F}_N$). Since $\mathcal{V}_{w_i}$ contains $\overline{\mathcal{V}}_{w_{2m-1}}$ for $i < 2m-1$ the connectedness follows.

To prove (4) consider the reduced expression for w_i for $i \leq m$: it is $s_i s_{i+1} \cdots s_m s_{m-1} \cdots s_1$, see [E-vdG2, Lemma 11.1]. This shows that all the s_i occur in it and we see that the $\overline{\mathcal{U}}_{s_i}$ for $i = 1, \ldots, m$ occur in the closure of $\overline{\mathcal{U}}_{w_i}$.

The proof of (2) is similar to the proof of [E-vdG1, Proposition 6.1] and uses induction on the Bruhat order. If $\ell(w) \leq 2m-2$ then $\overline{\mathcal{U}}_w$ is proper in $\mathcal{B}_N$. If an irreducible component has a non-empty intersection with a $\overline{\mathcal{U}}_{w'}$ with $w' > w$, then induction provides a point of $\mathcal{U}_1$; otherwise, we can apply a version of the Raynaud trick as in [E-vdG1, Lemma 6.2] and conclude that $w = 1$. If $\ell(w) = 2m-1$, then the image of any irreducible component Y of $\overline{\mathcal{U}}_w$ in $\mathcal{F}_N$ is either contained in $\overline{\mathcal{V}}_{w_3}$ and then Y is proper in $\mathcal{B}_N$ or the image coincides with $\overline{\mathcal{V}}_{w_2}$. In the latter case it maps in a generically finite way to it and therefore any irreducible component Y of $\overline{\mathcal{U}}_w$ intersects the fibers over the superspecial points, hence by induction contains a point of $\mathcal{U}_1$.

For (3) we now look in the fiber Z of the flag space over the image of a point of $\mathcal{U}_1$. This corresponds to a K3 surfaces for which the Hodge filtration $U_{-1} \subset U_0 \subset U_1 \subset U_2 = H$ coincides with the conjugate filtration $U_{-1}^c \subset U_0^c \subset U_1^c \subset U_2^c = H$. Moreover, we have the identifications

$$F^*(U_0) \cong (U_2^c/U_1^c) = (U_0^c)^\vee = U_0^\vee,$$

given by Cartier and the intersection pairing and similarly

$$F^*(U_1/U_0) \cong (U_1/U_0)^\vee,$$

giving U_0 and U_1/U_0 (and also U_2/U_1)) the structure of a p-unitary space. Indeed, if S is an $\mathbb{F}_p$-scheme then a p-unitary vector bundle $\mathcal{E}$ over S is a vector bundle together with an isomorphism $F^*(\mathcal{E}) \cong \mathcal{E}^*$ with F the absolute Frobenius. This gives rise to a bi-additive map $\langle\, ,\, \rangle : \mathcal{E} \times \mathcal{E} \to \mathcal{O}_S$ satisfying $\langle fx, y\rangle = f^p\langle x, y\rangle$ and $\langle x, fy\rangle = f\langle x, y\rangle$ for f a section of $\mathcal{O}_S$. In the étale topology this notion is equivalent to a local system of $\mathbb{F}_{p^2}$-vector spaces, cf. [E-vdG1, Proposition 7.2].

In case that $S = \mathrm{Spec}(\mathbb{F}_{p^2})$ we can then consider the flag space Z of complete p-unitary flags on U_1/U_0. The smallest p-unitary stratum there is the stratum of flags that coincide with their p-unitary dual. Such flags are defined over $\mathbb{F}_{p^2}$ as one sees by taking the dual once more.

So we look now at self-dual flags $A_\bullet = \{A_1 \subset A_2 \subset \cdots \subset A_{2m-1}\}$ on $A = U_1/U_0$ and we compare the flag $A_\bullet^{(p)}$ with the flag $A_\bullet$. (Note that the indices i now run from 1 to $m-1$ instead of from 1 to m since we leave U_0 fixed.) For an element $s = (i, i+1)$ with $1 \leq i \leq m-2$ we look at the intersection of the stratum $\overline{\mathcal{U}}_s$ with the fiber Z. It consists of those flags $A_\bullet$ with the property that the steps A_j for $j \neq i$ and $j \neq 2m-1-i$ are $\mathbb{F}_{p^2}$-rational and that for all j we have $A_j = A_{2m-1-j}^{\perp}$. We see that we can choose A_i freely by prescribing its image in A_{i+1}/A_{i-1}, hence this locus is a $\mathbb{P}^1$. In case $s = (m-1, m+1)$ we have to choose a space A_{m-1} and its orthogonal $A_m = A_{m-1}^{\perp}$ in $A_{m-1} \subset A_{m-1} \subset A_m \subset A_{m+1}$. Now all non-degenerate p-unitary forms are equivalent, so we may choose the form $x^{p+1} + y^{p+1} + z^{p+1}$ on the 3-dimensional space A_{m+1}/A_{m-2}. So in this case $\overline{\mathcal{U}}_s$ is isomorphic to the Fermat curve. The points of $\mathcal{U}_1$ are the $\mathbb{F}_{p^2}$-rational points on it.

In case the space $A = U_1/U_0$ is even-dimensional, say $\dim A = 2m-2$ the same argument works for s_i with $1 \leq i \leq m-2$. For $s = (m-2, m)(m-1, m+1)$ (resp. for $s = (m-1, m)$) we remark that it equals $s' s_{m-2} s'$ with $s' = (m-1, m)$, hence we find a $\mathbb{P}^1$ by picking a flag $A_{m-2} \subset A_{m-1} \subset A_m$.

This shows that we remain in the same connected component of $\overline{\mathcal{U}}^{(1)} \cap Z$, with $\mathcal{U}^{(1)}$ the union of the 1-dimensional strata $\mathcal{U}_v$, if we change the flag $A_\bullet$ at place i and $2m-1-i$ compatibly. By Lemma 7.6 of [E-vdG1] this implies that $\overline{\mathcal{U}}^{(1)} \cap Z$ is connected. □

References

[A] M. Artin, Supersingular *K*3 surfaces. Ann. Sci. École Norm. Sup. **7**, 543–567 (1974)

[A-M] M. Artin, B. Mazur, Formal groups arising from algebraic varieties. Ann. Sci. Ecole Norm. Sup. **10**, 87–132 (1977)

[A-SD] M. Artin, H.P.F. Swinnerton-Dyer, The Shafarevich-Tate conjecture for pencils of elliptic curves on *K*3 *surfaces*. Invent. Math. **20**, 249–266 (1973)

[C] F. Charles, The Tate conjecture for K3 surfaces over finite fields. Invent. Math. **194**(1), 119–145 (2013)

[C-O] C-L. Chai, F. Oort, Monodromy and irreducibility of leaves. Ann. Math. (2) **173**, 1359–1396 (2011)

[E] T. Ekedahl, On supersingular curves and abelian varieties. Math. Scand. **60**, 151–178 (1987)

[E-vdG1] T. Ekedahl, G. van der Geer, Cycle classes of the E-O stratification on the moduli of abelian varieties, in *Algebra, Arithmetic and Geometry*. Progress in Mathematics, vol. 269–270 (Birkhäuser, Boston, 2010)

[E-vdG2] T. Ekedahl, G. van der Geer, Cycle Classes on the Moduli of K3 surfaces in positive characteristic. Sel. Math. (N.S.) (2011, to appear). arXiv:1104.3024. http://dx.doi.org/10.1007/s00029-014-0156-8

[G-H-S] V. Gritsenko, K. Hulek, G. Sankaran, The Hirzebruch-Mumford volume for the orthognal group and applications. Doc. Math. **12**, 215–241 (2007)

[H] D. Huybrechts, Lectures on K3 surfaces. *Lecture Notes*. http://www.math.uni-bonn.de/people/huybrech/K3Global.pdf

[K-S] S. Kondo, I. Shimada, On certain duality of Néron-Severi lattices of supersingular K3 surfaces and its application to generic supersingular K3 surfaces. Algebraic Geom. **1**(3), 311–333 (2014)

[L] C. Liedtke, Supersingular k3 surfaces are unirational. Eprint. arXiv:1304.5623

[M] D. Maulik, Supersingular K3 surfaces for large primes. Eprint. arXiv:1203.2889

[Mo] B. Moonen, A dimension formula for Ekedahl-Oort strata. Ann. Inst. Fourier (Grenoble) **54**(3), 666–698 (2004)

[M-W] B. Moonen, T. Wedhorn, Discrete invariants of varieties in positive characteristic. Int. Math. Res. Not. **2004**(72), 3855–3903 (2004)

[Og1] A. Ogus, Supersingular K3 crystals. *Journées de Géométrie Algébrique de Rennes (II)*. Astérisque, vol. 64 (Société Mathématique de France, Paris, 1979), pp. 3–86.

[Og2] A. Ogus, Singularities of the height strata in the moduli of K3 surfaces, *Moduli of Abelian Varieties. (Texel Island, 1999)*. Progress in Mathematics, vol. 195 (Birkhäuser, Basel, 2001), pp. 325–343

[O] F. Oort, A stratification of a moduli space of abelian varieties, *Moduli of Abelian Varieties (Texel Island, 1999)*. Progress in Mathematics, vol. 195, (Birkhäuser, Basel, 2001) pp. 345–416

[P1] K. Madapusi Pera, Toroidal compactifications of integral models of Shimura varieties of Hodge type. Eprint. arXiv:1211.1731

[P2] K. Madapusi Pera, The Tate conjecture for K3 surfaces in odd characteristic. Eprint. arXiv:1301.6326

[P-R] H. Pittie, A. Ram, A Pieri-Chevalley formula in the K-theory of a G/B-bundle. Electron. Res. Announc. Am. Math. Soc. **5**, 102–107 (1999)

[R] M. Rapoport, On the Newton stratification. *Séminaire Bourbaki*. Astérisque No. 290, vol. 2001/2002, Exp. No. 903, viii, 207–224 (2003)

[R-S2] A.N. Rudakov, I.R. Shafarevich, Supersingular K3 surfaces over fields of characteristic 2. Izv. Akad. Nauk SSSR Ser. Mat. **42**, 848–869 (1978). Also in: I.R. Shafarevich: Collected Mathematical Papers. Springer Verlag.

[R-S] A.N. Rudakov, I.R. Shafarevich, Surfaces of type K3 over fields of finite characteristic. *Current Problems in Mathematics*, vol. 18 (Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Informatsii, Moscow, 1981), pp. 115–207. Also in: I.R. Shafarevich: Collected Mathematical Papers. Springer Verlag.

[R-S-Z] A.N. Rudakov, I.R. Shafarevich, Th. Zink, The effect of height on degenerations of K3 surfaces. Izv. Akad. Nauk SSSR Ser. Mat. **46**(1), 117–134, 192 (1982). Also in: I.R. Shafarevich: Collected Mathematical Papers. Springer Verlag.

[S] T. Shioda, Algebraic cycles on certain K3 surfaces in characteristic p. *Manifolds–Tokyo 1973 (Proc. Internat. Conf., Tokyo, 1973)* (University of Tokyo Press, Tokyo, 1975), pp. 357–364

[T] J. Tate, Algebraic cycles and poles of zeta functions. *Arithmetical Algebraic Geometry (Proc. Conf. Purdue Univ., 1963)* (Harper & Row, New York, 1965) pp. 93–110

[vdG] G. van der Geer, Cycles on the moduli space of abelian varieties. *Moduli of Curves and Abelian Varieties*. Aspects of Mathematics (vol. E33, pp. 65–89) (Vieweg, Braunschweig, 1999)

[vdG-K1] G. van der Geer, T. Katsura, On a stratification of the moduli of *K*3 surfaces. J. Eur. Math. Soc. (JEMS) **2**(3), 259–290 (2000)

[vdG-K2] G. van der Geer, T. Katsura, Note on tautological classes of moduli of K3 surfaces. Mosc. Math. J. **5**, 775–779 (2005)

The Right Adjoint of the Parabolic Induction

Marie-France Vignéras

Abstract We extend the results of Emerton on the ordinary part functor to the category of the smooth representations over a general commutative ring R, of a general reductive p-adic group G (rational points of a reductive connected group over a local non-archimedean field F of residual characteristic p). In Emerton's work, the characteristic of F is 0, R is a complete artinian local $\mathbb{Z}_p$-algebra having a finite residual field, and the representations are admissible. We show:

The smooth parabolic induction functor admits a right adjoint. The center-locally finite part of the smooth right adjoint is equal to the admissible right adjoint of the admissible parabolic induction functor when R is noetherian. The smooth and admissible parabolic induction functors are fully faithful when p is nilpotent in R.

1 Introduction

Let R be a commutative ring, let F be a local non-archimedean field of finite residual field of characteristic p, let $\mathbf{G}$ be a reductive connected F-group. Let $\mathbf{P}, \overline{\mathbf{P}}$ be two opposite parabolic F-subgroups of unipotent radicals $\mathbf{N}, \overline{\mathbf{N}}$ and same Levi subgroup $\mathbf{M} = \mathbf{P} \cap \overline{\mathbf{P}}$. Let $\mathbf{A_M}$ be the maximal F-split central subtorus of $\mathbf{M}$. The groups of F-points are denoted by the same letter but not in bold. The parabolic induction functor $\mathrm{Ind}_P^G : \mathrm{Mod}_R^\infty(M) \to \mathrm{Mod}_R^\infty(G)$ between the categories of smooth R-representations of M and of G, is the right adjoint of the N-coinvariant functor, and respects admissibility.

For any (R, F, G), we show that Ind_P^G admits a right adjoint R_P^G.

When R is noetherian, we show that the A_M-locally finite part of R_P^G respects admissibility, hence is the right adjoint of the functor Ind_P^G between admissible R-representations.

When 0 is the only infinitely p-divisible element in R, we show that the counit of the adjoint pair $(-_N, \mathrm{Ind}_P^G)$, is an isomorphism. Therefore, Ind_P^G is fully faithful and the unit of the adjoint pair $(\mathrm{Ind}_P^G, R_P^G)$ is an isomorphism.

M.-F. Vignéras (✉)
Institut de Mathematiques de Jussieu, Université de Paris 7, 4 rue Jussieu,
Paris 75252, France
e-mail: marie-france.vigneras@imj-prg.fr

W. Ballmann et al. (eds.), *Arbeitstagung Bonn 2013*, Progress in Mathematics 319,
DOI 10.1007/978-3-319-43648-7_15

The results of this paper have already be used in [HV] to compare the parabolic and compact inductions of smooth representations over an algebraically closed field R of characteristic p for any pair $(F, \mathbf{G})$, following the arguments of Herzig when the characteristic of F is 0 and $\mathbf{G}$ is split. The comparison is a basic step in the classification of the non-supersingular admissible irreducible representations of G (work in progress[1] with Abe, Henniart, and Herzig, see also Ly's work [Ly] for $GL(n, D)$ where D/F is a finite dimensional division algebra).

When p is invertible in R, it was known that Ind_P^G has a right adjoint, called also the "second adjoint." When R is the field of complex numbers, Casselman for admissible representations, and Bernstein in general proved that the right adjoint is equal to the $\overline{N}$-coinvariant functor multiplied by the modulus of P. Another proof was published by Bushnell [Bu]. Both proofs rely on the property that the category $\mathrm{Mod}_{\mathbb{C}}(G)$ is noetherian. Conversely, Dat [Dat] proved that the second adjointness implies the noetheriannity of $\mathrm{Mod}_R(G)$ and prove it assuming the existence of certain idempotents (constructed using the theory of types for linear groups, classical groups if $p \neq 2$, and groups of semi-simple rank 1). Under this hypothesis on G, Dat showed also that the N-coinvariant functor respects admissibility.

When the characteristic of F is 0 and R is a complete artinian local $\mathbb{Z}_p$-algebra having finite residual field, Emerton [Emerton] showed that Ind_P^G restricted to admissible representations has a right adjoint equal to the ordinary part functor $\mathrm{Ord}_{\overline{P}}$. Introducing the derived ordinary functors he showed also that the N-coinvariant functor respects admissibility [Emerton2, 3.6.7 Cor].

In Sect. 2 we give precise definitions and references to the literature on adjoint functors and on grothendieck abelian categories.

In Sects. 3 and 4, the existence of a right adjoint of $\mathrm{Ind}_P^G : \mathrm{Mod}_R^\infty(M) \to \mathrm{Mod}_R^\infty(G)$ is proved using that $\mathrm{Mod}_R^\infty(G)$ is a grothendieck abelian category and that Ind_P^G is an exact functor commuting with small direct sums. This method does not apply to the functor $\mathrm{Ind}_P^G : \mathrm{Mod}_R^{\mathrm{adm}}(M) \to \mathrm{Mod}_R^{\mathrm{adm}}(G)$ because the category of smooth admissible R-representations is not grothendieck in general. It is not even known if it is an abelian category when R is a field of characteristic p as well as F.

In Sect. 5, we assume that p is nilpotent in R; we show the vanishing of the N-coinvariants of ind_P^{PgP} when $PgP \neq P$ and that the counit of the adjunction $(-_N, \mathrm{Ind}_P^G)$ is an isomorphism; the general arguments of Sect. 2 imply that the unit of the adjunction $(\mathrm{Ind}_P^G, R_P^G)$ is an isomorphism and that Ind_P^G is fully faithful. When R is noetherian, $\mathrm{Ind}_P^G : \mathrm{Mod}_R^{\mathrm{adm}}(M) \to \mathrm{Mod}_R^{\mathrm{adm}}(G)$ is also obviously fully faithful.

In Sect. 6, we replace G by its open dense subset $P\overline{P}$. The partial compact induction functor $\mathrm{ind}_P^{P\overline{P}} : \mathrm{Mod}_R^\infty(M) \to \mathrm{Mod}_R^\infty(\overline{P})$ admits a right adjoint $R_P^{P\overline{P}}$ by the general method of Sect. 2. Let $\mathrm{Res}_{\overline{P}}^G : \mathrm{Mod}_R(G) \to \mathrm{Mod}_R(\overline{P})$ be the restriction functor. Let A_M be the split center of M. We fix an element $z \in A_M$ strictly contracting N. We prove that the z-locally finite parts of R_P^G and of $R_P^{P\overline{P}} \circ \mathrm{Res}_{\overline{P}}^G$ are

[1] This work is now written in N. Abe, Noriyuki, G. Henniart, F. Herzig, M.-F. Vignéras - A classification of admissible irreducible modulo p representations of reductive p-adic groups. To appear in Journal of the A.M.S. 2016.

isomorphic. The right adjoint $R_{\overline{P}}^{\overline{P}P} : \mathrm{Mod}_R^\infty(P) \to \mathrm{Mod}_R^\infty(M)$ of $\mathrm{ind}_{\overline{P}}^{\overline{P}P}$ is explicit: it is the smooth part of the functor $\mathrm{Hom}_{R[N]}(C_c^\infty(N,R), -)$.

In Sect. 7, following Casselman and Emerton, we give the Hecke description of the above functor $R_{\overline{P}}^{\overline{P}P} : \mathrm{Mod}_R^\infty P \to \mathrm{Mod}_R^\infty(M)$. We fix an open compact subgroup N_0 of N. The submonoid M^+ of elements of M contracting N_0 acts on V^{N_0} by the Hecke action. We have the smooth induction functor $\mathrm{Ind}_{M^+}^M : \mathrm{Mod}_R^\infty(M^+) \to \mathrm{Mod}_R^\infty(M)$. We show that $R_{\overline{P}}^{\overline{P}P}$ is the functor $V \mapsto \mathrm{Ind}_{M^+}^M(V^{N_0})$. The A_M-locally finite part of this functor is the Emerton's ordinary part functor $\mathrm{Ord}_P : \mathrm{Mod}_R^\infty P \to \mathrm{Mod}_R^\infty(M)$.

In Sect. 8 we assume that R is noetherian and we show that $\mathrm{Ord}_P(V)$ is admissible when V is an admissible R-representation of G. Therefore the parabolic induction functor $\mathrm{Ind}_P^G : \mathrm{Mod}_R^{\mathrm{adm}} M \to \mathrm{Mod}_R^{\mathrm{adm}} G$ admits a right adjoint equal to the functor $\mathrm{Ord}_{\overline{P}}^G : \mathrm{Ord}_{\overline{P}} \circ \mathrm{Res}_{\overline{P}}^G$. The unit of the adjunction $(\mathrm{Ind}_P^G, \mathrm{Ord}_{\overline{P}})$ is an isomorphism[2].

I thank Noriyuki Abe, Florian Herzig, Guy Henniart, and Michael Rapoport for their comments and questions, and the referee for an excellent report, allowing me to improve the paper and to correct some mistakes.

2 Review on Adjunction Between Grothendieck Abelian Categories

We fix a universe $\mathcal{U}$ and we denote by $\mathbb{S}et$ the category of $\mathcal{U}$-sets, i.e., belonging to $\mathcal{U}$. In a small category, the set of objects is $\mathcal{U}$-small, i.e., isomorphic to a $\mathcal{U}$-set, as well as the set of morphisms $\mathrm{Hom}(A, B)$ for any objects A and B. In a locally small category, only the set $\mathrm{Hom}(A, B)$ is supposed to be $\mathcal{U}$-small. (In [KS, 1.1, 1.2], small is called $\mathcal{U}$-small, and a locally small category is called a $\mathcal{U}$-category.)

Let $\mathcal{I}$ be a small category and let $\mathcal{C}, \mathcal{D}$ be locally small categories. We denote by $\mathcal{C}^{\mathrm{op}}$ the opposite category of $\mathcal{C}$ and by $\mathcal{D}^{\mathcal{C}}$ the category of functors $\mathcal{C} \to \mathcal{D}$. A contravariant functor $\mathcal{C} \to \mathcal{D}$ is a functor $\mathcal{C}^{\mathrm{op}} \to \mathcal{D}$. The categories $\mathbb{S}et^{\mathcal{C}^{\mathrm{op}}}, \mathbb{S}et^{\mathcal{C}}$ are not locally small in general (if $\mathcal{C}$ is not small) [KS, Definition 1.4.2].

Proposition 2.1 ([KS, Definition 1.2.11, Corollary 1.4.4]). *The contravariant Yoneda functor : $C \mapsto \mathrm{Hom}(C, -) : \mathcal{C} \to \mathbb{S}et^{\mathcal{C}}$ and the covariant Yoneda functor : $C \mapsto \mathrm{Hom}(-, C) : \mathcal{C} \to \mathbb{S}et^{\mathcal{C}^{\mathrm{op}}}$ are fully faithful.*

A functor F in $\mathbb{S}et^{\mathcal{C}}$ or in $\mathbb{S}et^{\mathcal{C}^{\mathrm{op}}}$ is called representable when it is isomorphic to the image of an object $C \in \mathcal{C}$ by the Yoneda functor [KS, Definition 1.4.8]. The object C which is unique modulo unique isomorphism is called a representative of F.

[2] We know now that R_P^G respects the admissibility (N. Abe, Noriyuki, G. Henniart, F. Herzig, M.-F. Vignéras - Mod p representations of reductive p-adic groups: functorial properties. In progress, 2016.)

A functor $F : \mathcal{I} \to \mathcal{C}$ defines functors

$$\varinjlim F \in \mathbb{S}et^{\mathcal{C}} \quad C \mapsto \mathrm{Hom}_{\mathcal{C}^{\mathcal{I}}}(F, ct_C), \quad \varprojlim F \in \mathbb{S}et^{\mathcal{C}^{\mathrm{op}}} \quad C \mapsto \mathrm{Hom}_{\mathcal{C}^{\mathcal{I}}}(ct_C, F),$$

where $ct_C : \mathcal{I} \to \mathcal{C}$ is the constant functor defined by $C \in \mathcal{C}$. When the functor $\varinjlim F$ is representable, a representative is called the injective limit (or colimit or direct limit) of F, is denoted also by $\varinjlim F$, and we have natural isomorphism [ML, III.4 (2), (3)]

$$\varinjlim F(C) = \mathrm{Hom}_{\mathcal{C}^{\mathcal{I}}}(F, ct_C) \simeq \mathrm{Hom}_{\mathcal{C}}(\varinjlim F, C).$$

When the functor $\varprojlim F$ is representable, a representative is called the projective limit (or inverse limit or limit) of F, is denoted also by $\varprojlim F$, and we have natural isomorphism

$$\varprojlim F(C) = \mathrm{Hom}_{\mathcal{C}^{\mathcal{I}}}(ct_C, F) \simeq \mathrm{Hom}_{\mathcal{C}}(C, \varprojlim F).$$

One says also that $(F(i))_{i\in\mathcal{I}}$ is an inductive or projective system in $\mathcal{C}$ indexed by $\mathcal{I}$ or $\mathcal{I}^{\mathrm{op}}$ and one writes $\varinjlim(F(i))_{i\in\mathcal{I}}$ or $\varprojlim(F(i))_{i\in\mathcal{I}^{\mathrm{op}}}$ for the object $\varinjlim F$ or $\varprojlim F$.

Example 2.2. (1) A set of objects $(C_i)_{i\in\mathcal{I}}$ of $\mathcal{C}$ indexed by a set $\mathcal{I}$ can be viewed as a functor $F : \mathcal{I} \to \mathcal{C}$ where $\mathcal{I}$ is identified with a discrete category (the only morphisms are the identities). When they exist, $\varinjlim F = \oplus_{i\in\mathcal{I}} C_i$ is the direct sum, or coproduct, or disjoint union $\sqcup_{i\in\mathcal{I}} C_i$, and $\varprojlim F = \prod_{i\in\mathcal{I}} C_i$ is the direct product.

(2) When $\mathcal{I}$ has two objects and two parallel morphisms other than the identities, a functor $F : \mathcal{I} \to \mathcal{C}$ is nothing but two parallels arrows $C_1 \underset{f}{\overset{g}{\rightrightarrows}} C_2$ in $\mathcal{C}$. When they are representable, $\varinjlim F$ is the cokernel of (f, g) and $\varprojlim F$ is its kernel [KS, Definition 2.2.2].

(3) When they are representable, it is possible to construct the inductive (resp. projective) limit of a functor $F : \mathcal{I} \mapsto \mathcal{C}$, using only coproduct and cokernels (resp. products and kernels) [KS, Proposition 2.2.9]. If $\mathrm{Hom}(\mathcal{I})$ denotes the set of morphisms $s : \sigma(s) \to \tau(s)$ with $\sigma(s), \tau(s) \in \mathcal{I}$, of the category $\mathcal{I}$,

$$\varinjlim F \text{ is the cokernel of } f, g : \oplus_{s\in\mathrm{Hom}(\mathcal{I})} F(\sigma(s)) \underset{f}{\overset{g}{\rightrightarrows}} \oplus_{i\in\mathcal{I}} F(i), \tag{1}$$

where f, g correspond, respectively, to the two morphisms $\mathrm{id}_{F(\sigma(s))}, F(s)$, for $s \in \mathrm{Hom}(\mathcal{I})$,

$$\varprojlim F \text{ is the kernel of } \prod_{i\in\mathcal{I}} F(i) \underset{f}{\overset{g}{\rightrightarrows}} \prod_{s\in\mathrm{Hom}(\mathcal{I})} F(\sigma(s)),$$

where f, g are deduced from the morphisms $\mathrm{id}_{F(\tau(s))}, F(s) : F(\tau(s)) \times F(\sigma(s)) \underset{f}{\overset{g}{\rightrightarrows}} F(\tau(s))$ for $s \in \mathrm{Hom}(\mathcal{I})$.

A non-empty category $\mathcal{C}$ is called filtrant if, for any two objects C_1, C_2 there exist morphisms $C_1 \to C_3, C_2 \to C_3$, and for any parallel morphisms $C_1 \overset{g}{\underset{f}{\rightrightarrows}} C_2$, there exists a morphism $h : C_2 \to C_3$ such that $h \circ f = h \circ g$ [KS, Definition 3.1.1].

Let $F : \mathcal{C} \mapsto \mathcal{D}$ be a functor. For $U \in \mathcal{D}$, we have the category $\mathcal{C}_U$ whose objects are the pairs (X, u) with $X \in \mathcal{C}, u : F(X) \to U$ in $\mathrm{Hom}(\mathcal{D})$. We say that F is right exact if the category $\mathcal{C}_U$ is filtrant for any $U \in \mathcal{D}$, and that F is left exact if the functor $F^{\mathrm{op}} : \mathcal{D}^{\mathrm{op}} \to \mathcal{C}^{\mathrm{op}}$ is right exact [KS, 3.3.1].

Proposition 2.3. *Let a functor $F : \mathcal{C} \mapsto \mathcal{D}$.*

(1) When $\mathcal{C}$ admits finite projective limits, F is left exact if and only if it commutes with finite projective limits. In this case, F commutes with the kernel of parallel arrows.
(2) When $\mathcal{C}$ admits small projective limits, F is left exact and commutes with small direct products, if and only if F commutes with small projective limits.
(3) The similar statements hold true for right exact functors, inductive limits, small direct sums, and cokernels.

Proof. (1) See [KS, Proposition 3.3.3, Corollary 3.3.4].
(2) If F preserves small projective limits, F is left exact and preserves small direct products [Example 2.2 (1)]. Conversely, from (1), a left exact functor which commutes with small direct products preserves small projective limits because it commutes with the kernel of the parallel arrows.
(3) Replace $\mathcal{C}$ by $\mathcal{C}^{\mathrm{op}}$. ☐

Let $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{C}$ be two functors. Then (F, G) is a pair of adjoint functors, or F is the left adjoint of G, or G is the right adjoint if F, if their exists an isomorphism of bifunctors from $\mathcal{C}^{\mathrm{op}} \times \mathcal{C}$ to $\mathbb{S}et$

$$\mathrm{Hom}_{\mathcal{D}}(F(.), .) \simeq \mathrm{Hom}_{\mathcal{C}}(., G(.)),$$

called the adjunction isomorphism [KS, Definition 1.5.2]. The functor F determines the functor G up to unique isomorphism and G determines F up to unique isomorphism [KS, Theorem 1.5.3]. For $X \in \mathcal{C}$, the image of the identity $\mathrm{id}_{F(X)} \in \mathrm{Hom}_{\mathcal{D}}(F(X), F(X))$ by the adjunction isomorphism is a morphism $X \mapsto G \circ F(X)$. Similarly, for $Y \in \mathcal{D}$, the image of $\mathrm{id}_{G(Y)}$ is a morphism $F \circ G(Y) \to Y$. The morphisms are functorial in X and Y. The corresponding morphisms of functors are called the unit and the counit :

$$\epsilon : 1_{\mathcal{C}} \to G \circ F, \quad \eta : F \circ G \to 1_{\mathcal{D}}.$$

Proposition 2.4. *Let (F, G) be a pair of adjoint functors.*
F is fully faithful if and only if the unit $\epsilon : 1 \to G \circ F$ is an isomorphism.
G is fully faithful if and only if the counit $\eta : F \circ G \to 1$ is an isomorphism.

F and G are fully faithful if and only if F is an equivalence (fully faithful and essentially surjective [KS, Definition 1.2.11,1.3.13] *if and only if G is an equivalence. In this case F and G are quasi-inverse one to each other.*

Proof. See [KS, Proposition 1.5.6]. □

Proposition 2.5. *Let (F, G) be a pair of adjoint functors. Then F is right exact and G is left exact.*

Proof. See [KS, Proposition 3.3.6]. □

Let $\mathcal{A}$ be a locally small abelian category. A generator of $\mathcal{A}$ is an object $E \in \mathcal{A}$ such that the functor $\mathrm{Hom}(E, -) : \mathcal{A} \to \mathbb{S}et$ is faithful (i.e., any object of $\mathcal{A}$ is a quotient of a small direct sum $\oplus_i E$). If $\mathcal{A}$ admits small inductive limits, the functor between abelian categories

$$F \mapsto \varinjlim F : \mathcal{A}^{\mathcal{I}} \to \mathcal{A}$$

is additive and right exact.

Definition 2.6 ([KS, Definition 8.3.24]). A locally small abelian category $\mathcal{A}$ is called grothendieck if it admits a generator, small inductive limits, and the small filtered inductive limits are exact.

Example 2.7. Given a ring $R \in \mathcal{U}$, the category of left R-modules in $\mathcal{U}$ is small, abelian, and grothendieck with generator R.

Proof. See [KS, Example 8.3.25]. □

Proposition 2.8. *A grothendieck abelian locally small category admits small projective limits.*

Proof. See [KS, Proposition 8.3.27]. □

Proposition 2.9. *Let a functor $F : \mathcal{A} \to \mathcal{C}$ where $\mathcal{A}$ is a grothendieck abelian locally small category. The following properties are equivalent:*

(1) F admits a right adjoint,
(2) F commutes with small inductive limits,
(3) F is right exact and commutes with small direct sums.

Proof. See [KS, Proposition 8.3.27]. □

A similar statement characterizes the existence of a left adjoint.

3 The Category $\mathrm{Mod}_R^\infty(G)$

Let R be a commutative ring, let G be a second countable locally profinite group (for instance, a parabolic subgroup of a reductive p-adic group), and let $(K_n)_{n\in\mathbb{N}}$ be a strictly decreasing sequence of pro-p-open subgroups of G, with trivial intersection, such that K_n normal in K_0 for all n.

3.1 $\mathrm{Mod}_R^\infty(G)$ is Grothendieck

A R-representation V of G is a left $R[G]$-module. A vector $v \in V$ is called smooth when it is fixed by an open subgroup of G. The set of smooth vectors of V is a $R[G]$-submodule of V, equal to $V^\infty = \cup_{n\in\mathbb{N}} V^{K_n}$ where V^{K_n} is the submodule of $v \in V$ fixed by K_n. When every vector of V is smooth, V is called smooth. [The same definition applies to a locally profinite monoid (the maximal subgroup is open and locally profinite).]

Example 3.1. The module $C_c(G, R)$ of functions $f : G \to R$ with compact support is a $R[G \times G]$-module for the left and right translations. For $n \in \mathbb{N}$, the submodule $C_c(K_n\backslash G, R)$ of compactly supported functions left invariant by K_n is a smooth representation of G for the right translation. These submodules form a strictly increasing sequence of union the smooth part $C_c^\infty(G, R)$ of $C_c(G, R)$.

We allow only the R-modules of cardinal $< c$ for some uncountable strong limit cardinal $c > |R|$, so that the R-representations of G form a set rather than a proper class (we work in the same artinian universe $\mathcal{U}_c$ [SGA4, Exposé 1, p. 4]; the cardinal of $\mathrm{Hom}_{RG}(V, V')$ is $< c$ for two R-representations V, V' of G). The abelian category $\mathrm{Mod}_R(G)$ of left $R[G]$-modules is small, grothendieck of generator $R[G]$ (Example 2.7), and contains the abelian full subcategory $\mathrm{Mod}_R^\infty(G)$ of smooth R-representations of G.

Lemma 3.2. $\mathrm{Mod}_R^\infty(G)$ *is a grothendieck category of generator* $\oplus_{n\in\mathbb{N}} C_c(K_n\backslash G, R)$.

Proof. An arbitrary direct sum of smooth R-representations of G is smooth. The cokernel of two parallel arrows in $\mathrm{Mod}_R^\infty(G)$ is smooth hence $\mathrm{Mod}_R^\infty(G)$ admits small inductive limits [Example 2.2 (3)]. Small filtered inductive limits are exact because they are already exact in the grothendieck category $\mathrm{Mod}_R(G)$. □

For $W \in \mathrm{Mod}_R^\infty(G)$, $V \in \mathrm{Mod}_R(G)$ we have $\mathrm{Hom}_{R[G]}(W, V) = \mathrm{Hom}_{R[G]}(W, V^\infty)$. The smoothification

$$V \mapsto V^\infty : \mathrm{Mod}_R(G) \to \mathrm{Mod}_R^\infty(G)$$

is the right adjoint of the inclusion $\mathrm{Mod}_R^\infty(G) \to \mathrm{Mod}_R(G)$, hence is left exact (Proposition 2.5). The smoothification is never right exact if G is not the trivial group [Viglivre, I.4.3] hence does not have a right adjoint (Proposition 2.5).

3.2 Admissibility and z-Finiteness

Definition 3.3. An R-representation V of G is called admissible when it is smooth and for any compact open subgroup H of G, the R-module V^H of H-fixed elements of V is finitely generated.

When R is a noetherian ring, we consider the category $\mathrm{Mod}_R^{\mathrm{adm}}(G)$. It may not have a generator or small inductive limits. Worse, it may be not abelian.

Example 3.4. Let R be an algebraically closed field of characteristic p and $G = \mathbf{G}(F)$ a group as in the introduction. Given an open pro-p-subgroup I of G, a non-zero smooth R-representation of G contains a non-zero vector fixed by I; the set of irreducible admissible R-representations of G (modulo isomorphism) is infinite. Therefore their direct sum is not admissible. But it is a quotient of a generator of $\mathrm{Mod}_R^{\mathrm{adm}}(G)$, if a generator exists. If the quotient of an admissible representation remains admissible, a generator cannot exist. The admissibility is preserved by quotient when the characteristic of F is zero [VigLang], but this is unknown when the characteristic of F is p.

Let H any subset of the center of G, and let $V \in \mathrm{Mod}_R(G)$.

Definition 3.5. An element $v \in V$ is called H-finite if the R-module $R[H]v$ is contained in a finitely generated R-submodule of V.

The subset $V^{H\text{-lf}}$ of H-finite elements is a R-subrepresentation of V, called the H-locally finite part of V. When every element of V is H-finite, V is called H-locally finite. The category $\mathrm{Mod}_R^{H\text{-lf}}(G)$ of H-locally finite smooth R-representations of G is a full abelian subcategory of $\mathrm{Mod}_R^\infty(G)$. The H-locally finite functor

$$V \mapsto V^{H\text{-lf}} : \mathrm{Mod}_R^\infty(G) \to \mathrm{Mod}_R^{H\text{-lf}}(G) \tag{2}$$

is the right adjoint of the inclusion $\mathrm{Mod}_R^{H\text{-lf}}(G) \to \mathrm{Mod}_R^\infty(G)$.

Lemma 3.6. *If V is admissible, then V is H-locally finite.*

Proof. Let $v \in V$. As V is smooth, $v \in V^{K_n}$ for some $n \in \mathbb{N}$. As V is admissible, V^{K_n} is a finitely generated R-module. As H is central, V^{K_n} is H-stable. □

4 The Right Adjoint R_P^G of $\mathrm{Ind}_P^G : \mathrm{Mod}_R^\infty(M) \to \mathrm{Mod}_R^\infty(G)$

Let F be a local non-archimedean field of finite residue field of characteristic p, let $\mathbf{G}$ be a reductive connected F-group. We fix a maximal F-split subtorus $\mathbf{S}$ of $\mathbf{G}$, and a minimal parabolic F-subgroup $\mathbf{B}$ of $\mathbf{G}$ containing $\mathbf{S}$. We suppose that $\mathbf{S}$ is not trivial. Let $\mathbf{U}$ be the unipotent radical of $\mathbf{B}$. The $\mathbf{G}$-centralizer $\mathbf{Z}$ of $\mathbf{S}$ is a Levi subgroup of $\mathbf{B}$. We choose a pair of opposite parabolic F-subgroups $\mathbf{P}, \overline{\mathbf{P}}$ of $\mathbf{G}$ with $\mathbf{P}$ containing $\mathbf{B}$, of unipotent radicals $\mathbf{N}, \overline{\mathbf{N}}$ and Levi subgroup $\mathbf{M} = \mathbf{P} \cap \overline{\mathbf{P}}$. Let $\mathbf{A_M} \subset \mathbf{S}$ be the maximal F-split central subtorus of $\mathbf{M}$. We denote by X the group of F-rational points of an algebraic group $\mathbf{X}$ over F, with the exception that we write $N_G(S)$ for the group of F-rational points of the $\mathbf{G}$-normalizer $N_{\mathbf{G}}(\mathbf{S})$ of $\mathbf{S}$. The finite Weyl group is $W_0 = N_{\mathbf{G}}(\mathbf{S})/\mathbf{Z} = N_G(S)/Z$. We fix a strictly decreasing sequence $(K_n)_{n\in\mathbb{N}}$ of pro-p-open subgroups of G with trivial intersection, such that for all n, K_n is normal in K_0 and has an Iwahori decomposition

$$K_n = \overline{N}_n M_n N_n = N_n M_n \overline{N}_n, \tag{3}$$

where $M_n := K_n \cap M, N_n := K_n \cap N, \overline{N}_n := K_n \cap \overline{N}$.

For $W \in \mathrm{Mod}_R^\infty(M)$, the representation $\mathrm{Ind}_P^G(W) \in \mathrm{Mod}_R^\infty(G)$ smoothly parabolically induced by W is the R-module of functions $f : G \to W$ such that $f(mngx) = mf(g)$ for $m \in M, n \in N, g \in G, x \in K_n$ where $n \in \mathbb{N}$ depends on f, with G acting by right translations. The smooth parabolic induction

$$\mathrm{Ind}_P^G : \mathrm{Mod}_R^\infty(M) \to \mathrm{Mod}_R^\infty(G)$$

is the right adjoint of the N-coinvariant functor [Viglivre, I.5.7 (i), I.A.3 Proposition]

$$V \mapsto V_N \ : \ \mathrm{Mod}_R^\infty(G) \to \mathrm{Mod}_R^\infty(M) \ .$$

The N-coinvariant functor $\mathrm{Mod}_R(P) \to \mathrm{Mod}_R(M)$ is the left adjoint of the inflation functor $\mathrm{Infl}_M^P : \mathrm{Mod}_R(M) \to \mathrm{Mod}_R(P)$ sending a representation of $M = P/N$ to the natural representation of P trivial on N.

Remark 4.1. The N-coinvariants of the inflation functor Infl_M^P is the identity functor of $\mathrm{Mod}_R M$ (the counit $-_N \circ \mathrm{Infl}_M^P \to 1$ of the adjunction $(-_N, \mathrm{Infl}_M^P)$ is an isomorphism).

Proposition 4.2. *The smooth parabolic induction functor* $\mathrm{Ind}_P^G : \mathrm{Mod}_R^\infty(M) \to \mathrm{Mod}_R^\infty(G)$ *is exact, and admits a right adjoint*

$$R_P^G : \mathrm{Mod}_R^\infty(G) \to \mathrm{Mod}_R^\infty(M).$$

Proof. For $W \in \mathrm{Mod}_R^\infty(M)$, we write $C^\infty(P\backslash G, W)$ for the R-module of locally constant functions on the compact set $P\backslash G$ with values in W. We fix a continuous section

$$\varphi : P\backslash G \to G. \tag{4}$$

The R-linear map

$$f \mapsto f \circ \varphi : \mathrm{Ind}_P^G(W) \to C^\infty(P\backslash G, W) \tag{5}$$

is an isomorphism. We have a natural isomorphism

$$C^\infty(P\backslash G, W) \simeq C^\infty(P\backslash G, R) \otimes_R W \simeq C^\infty(P\backslash G, \mathbb{Z}) \otimes_{\mathbb{Z}} W. \tag{6}$$

The $\mathbb{Z}$-module $C^\infty(P\backslash G, \mathbb{Z})$ is free, because it is the union of the increasing sequence of the $\mathbb{Z}$-modules $L_n := C^\infty(P\backslash G/K_n, \mathbb{Z})$ for $n \in \mathbb{N}$, which are free of finite rank as well as the quotients L_n/L_{n+1}. Hence the tensor product by $C^\infty(P\backslash G, \mathbb{Z})$ is exact, and Ind_P^G is also exact.

The smooth parabolic induction commutes with small direct sums $\oplus_{i\in\mathcal{I}} W_i$ because a function $f \in C^\infty(P\backslash G, W)$ takes only finitely many values.

Applying Proposition 2.9 and Lemma 3.2, the parabolic induction admits a right adjoint. ◻

Remark 4.3. When p is invertible in R, Dat [Dat, between Corollary 3.7 and Proposition 3.8] showed that

$$R_P^G(V) = ([\mathrm{Hom}_{R[G]}(C_c^\infty(G,R),V)]^N)^\infty \quad (V \in \mathrm{Mod}_R^\infty(G)).$$

The modulus δ_P of P is well defined. When R is the field of complex numbers (Bernstein) or when G is a linear group, a classical group when $p \neq 2$, or of semi-simple rank 1 [Dat], we have:

$$R_P^G(V) \simeq \delta_P V_{\overline{N}}.$$

Let $g \in G$ and Q an arbitrary closed subgroup of G. The partial compact smooth parabolic induction functor

$$\mathrm{ind}_P^{PgQ} : \mathrm{Mod}_R^\infty(M) \to \mathrm{Mod}_R^\infty(Q)$$

associates to $W \in \mathrm{Mod}_R^\infty(M)$ the smooth representation $\mathrm{ind}_P^{PgQ}(W)$ of Q by right translation on the module of functions $f : PgQ \to W$ with compact support modulo left multiplication by P ($P\backslash PgQ$ is generally not closed in the compact set $P\backslash G$) such that $f(mnghx) = mf(gh)$ for $m \in M, n \in N, h \in Q, x \in K_n \cap Q$ where $n \in \mathbb{N}$ depends on f.

Remark 4.4. When $PgP = P$, the functor $\mathrm{ind}_P^P : \mathrm{Mod}_R^\infty(M) \to \mathrm{Mod}_R^\infty(P)$ is the inflation functor Infl_M^P.

Proposition 4.5. *The functor* ind_P^{PgQ} *is exact, and admits a right adjoint*

$$R_P^{PgQ} : \mathrm{Mod}_R^\infty(Q) \to \mathrm{Mod}_R^\infty(M).$$

Proof. Same proof as for the functor Ind_P^G (Proposition 4.2). □

Lemma 4.6. $W \in \mathrm{Mod}_R^\infty(M)$ *is admissible if and only if* $\mathrm{Ind}_P^G(W) \in \mathrm{Mod}_R^\infty(G)$ *is admissible.*

Proof. This is well known and follows from the decomposition [Viglivre, I.5.6, II.2.1]:

$$(\mathrm{Ind}_P^G W)^{K_n} \simeq \oplus_{PgK_n}(\mathrm{Ind}_P^{PgK_n} W)^{K_n} \simeq \oplus_{PgK_n} W^{M\cap gK_ng^{-1}} \quad (n \in \mathbb{N}, g \in G),$$

where the sum is finite and $\mathrm{Ind}_P^{PgK_n} W \subset \mathrm{Ind}_P^G W$ is the R-submodule of functions with support contained in PgK_n. □

Corollary 4.7. *When the ring R is noetherian, the smooth parabolic induction restricts to a functor, called the admissible parabolic induction,*

$$\mathrm{Ind}_P^G : \mathrm{Mod}_R^{\mathrm{adm}}(M) \to \mathrm{Mod}_R^{\mathrm{adm}}(G).$$

We will later show that the admissible parabolic induction admits also a right adjoint.

5 Ind_P^G is Fully Faithful if p is nilpotent in R

We keep the notation of the preceding section. Let Φ_G be the set of roots of S in G. We write U_α for the subgroup of G associated with a root $\alpha \in \Phi_G$ (the group $U_{(\alpha)}$ in [Bo, 21.9]).

Definition 5.1. The p-ordinary part R_{p-ord} of R is the subset of $x \in R$ which are infinitely p-divisible.

By [Viglivre, I (2.3.1)], $R_{p-ord} = \{0\}$ if and only if there exists no Haar measure on U_α with values in R. But p is nilpotent in R if and only if $R[1/p] = \{0\}$ if and only if

$$C_c^\infty(U_\alpha, R)_{U_\alpha} = 0. \tag{7}$$

When R is a field, $R_{p-ord} = \{0\}$ if and only if p is nilpotent in R if and only if the characteristic of R is $\neq p$.

Proposition 5.2. *We suppose that p is nilpotent in R. Let $W \in \mathrm{Mod}_R^\infty M$ and $g \in G$. The N-coinvariants of $\mathrm{ind}_P^{PgP}(W)$ is 0 if $PgP \neq P$.*

Proof. We identify $\mathrm{ind}_P^{PgP}(W)$ with $C_c^\infty(P\backslash PgP, R) \otimes_R W$ as in (5). The action of N on $C_c^\infty(P\backslash PgP, R) \otimes_R W$ is trivial on W and is the right translation on $C_c^\infty(P\backslash PgP, R)$. Therefore

$$(\mathrm{ind}_P^{PgP}(W))_N = C_c^\infty(P\backslash PgP, R)_N \otimes_R W,$$

and we can forget W. To show that $C_c^\infty(P\backslash PgP, R)_N = 0$ if $PgP \neq P$, we prove that there exists a B-positive root α such that $U_\alpha \subset N$ and the space $P\backslash PgP$ is of the form $X \times U_\alpha$ where the right action of U_α on $P\backslash PgP$ is trivial on X and equals the natural right action on U_α. Therefore

$$C_c^\infty(P\backslash PgP, R)_{U_\alpha} = C_c^\infty(X, R) \otimes_R C_c^\infty(U_\alpha, R)_{U_\alpha}.$$

Applying (7), we obtain $C_c^\infty(P\backslash PgP)_{U_\alpha} = 0$ hence $C_c^\infty(P\backslash PgP, R)_N = 0$.

It remains to explain the existence of such an α. As $(B, N_G(S))$ is a Tits system in G [BT1, 1.2.6], we have $PgP = P\nu P$ for an element $\nu \in N_G(S)$; we can suppose that the image w of ν in W_0 has minimal length in the double coset $W_{0,M} w W_{0,M}$ (where $W_{0,M} := N_M(S)/Z$). This implies that the fixator $N_\nu := \{n \in N \mid P\nu n = P\nu\}$ of $P\nu$ in N is generated by the U_α for the roots $\alpha \in \Phi_G - \Phi_M$ such that α and $w(\alpha)$ are reduced, B-positive. The fixator of $P\nu$ in M is a parabolic subgroup Q and the fixator of $P\nu$ in P is QN_ν. The group N is directly spanned by the U_β ($\beta \in \Phi_G - \Phi_M$ positive and reduced) taken in any order [Bo, 21.12]. As $PgP \neq P$, i.e., $w \neq 1$, there exists a reduced positive root $\alpha \in \Phi_G - \Phi_M$ such that $U_\alpha \not\subset N_\nu$. Such an α satisfies all the properties that we want. □

Theorem 5.3. *We suppose that p is nilpotent in R. Then*

1. *The parabolic induction* $\mathrm{Ind}_P^G : \mathrm{Mod}_R^\infty(M) \to \mathrm{Mod}_R^\infty(G)$ *is fully faithful,*
2. *The unit* $\mathrm{id}_{\mathrm{Mod}_R^\infty(M)} \to R_P^G \circ \mathrm{Ind}_P^G$ *of the adjoint pair* $(\mathrm{Ind}_P^G, R_P^G)$ *is an isomorphism.*
3. *The counit* $\eta : -_N \circ \mathrm{Ind}_P^G \to \mathrm{id}_{\mathrm{Mod}_R^\infty(M)}$ *of the adjoint pair* $(-_N, \mathrm{Ind}_P^G)$ *is an isomorphism.*

Proof. By Lemma 3.2 and Proposition 2.4, the three properties are equivalent. We prove that the counit η of the adjoint pair $(-_N, \mathrm{Ind}_P^G)$ is an isomorphism.

(a) It is well known that Ind_P^G admits a finite filtration $F_1 \subset \ldots \subset F_r$ of quotients ind_P^{PgP}, with last quotient ind_P^P, associated with $P\backslash G/P$.
(b) Being a right adjoint, the N-coinvariant functor $\mathrm{Mod}_R^\infty(P) \to \mathrm{Mod}_R^\infty(M)$ is right exact.
(c) Apply Proposition 5.2 and Remarks 4.1, 4.4. □

6 z-Locally Finite Parts of R_P^G and of $R_P^{P\overline{P}} \circ \mathrm{Res}_{\overline{P}}^G$ are Equal

We keep the notation of the preceding section. We fix an element $z \in A_M$ strictly contracting N : the sequence $(z^n N_0 z^{-n})_{n\in\mathbb{Z}}$ is strictly decreasing of trivial intersection and union N. We denote $N_n := z^n N_0 z^{-n}$ when $n < 0$ (N_n for $n \geq 0$ is defined in Sect. 4).

We compare the right adjoint $R_P^G : \mathrm{Mod}_R^\infty(G) \to \mathrm{Mod}_R^\infty(M)$ of the parabolic induction Ind_P^G to the functor $R_P^{P\overline{P}} \circ \mathrm{Res}_{\overline{P}}^G$, where $\mathrm{Res}_{\overline{P}}^G : \mathrm{Mod}_R^\infty(G) \to \mathrm{Mod}_R^\infty \overline{P}$ is the restriction functor and $R_P^{P\overline{P}} : \mathrm{Mod}_R^\infty(\overline{P}) \to \mathrm{Mod}_R^\infty(M)$ is the right adjoint of the partial compact parabolic induction $\mathrm{ind}_P^{P\overline{P}}$. We denote by

$$R_P^{G,z\text{-lf}} : \mathrm{Mod}_R^\infty G \to \mathrm{Mod}_R^{z\text{-lf}} M, \quad R_P^{P\overline{P},z\text{-lf}} : \mathrm{Mod}_R^\infty \overline{P} \to \mathrm{Mod}_R^{z\text{-lf}} M,$$

the z-locally finite parts of R_P^G and of $R_P^{P\overline{P}}$.

Theorem 6.1. *The functors* $R_P^{G,z\text{-lf}}$ *and* $R_P^{P\overline{P},z\text{-lf}} \circ \mathrm{Res}_{\overline{P}}^G$ *are isomorphic.*

Proof. We want to prove that there exists an isomorphism

$$\mathrm{Hom}_{R[M]}(W, R_P^{G,z\text{-lf}}(V)) \to \mathrm{Hom}_{R[M]}(W, R_P^{P\overline{P},z\text{-lf}}(V)) \tag{8}$$

functorial in $(W, V) \in \mathrm{Mod}_R^{z\text{-lf}}(M) \times \mathrm{Mod}_R^\infty(G)$. We may replace $R_P^{G,z\text{-lf}}, R_P^{P\overline{P},z\text{-lf}}$ by $R_P^G, R_P^{P\overline{P}}$ in (8) (recall (2)). Then using the adjunctions $(\mathrm{Ind}_P^G, R_P^G)$ and $(\mathrm{ind}_P^{P\overline{P}}, R_P^{P\overline{P}})$, we reduce to find an isomorphism

$$\mathrm{Hom}_{R[G]}(\mathrm{Ind}_P^G W, V) \to \mathrm{Hom}_{R[\overline{P}]}(\mathrm{ind}_P^{P\overline{P}} W, V) \tag{9}$$

functorial in $(W, V) \in \mathrm{Mod}_R^{z\text{-lf}}(M) \times \mathrm{Mod}_R^\infty(G)$. There is an obvious functorial homomorphism because $\mathrm{ind}_P^{P\overline{P}} W$ is a submodule of $\mathrm{Ind}_P^G W$. This homomorphism, denoted by J, sends a $R[G]$-homomorphism $\mathrm{Ind}_P^G W \to V$ to its restriction to $\mathrm{ind}_P^{P\overline{P}} W$. The homomorphism J is injective because an arbitrary open subset of $P\backslash G$ is a finite disjoint union of G-translates of compact open subsets of $P\backslash P\overline{P}$ [SVZ, Proposition 5.3]. To show that J is surjective, we introduce more notations.

Let $(g, r, \overline{n}, w) \in G \times \mathbb{N} \times \overline{N} \times W$. We say that $(g, r, \overline{n}, w)$ is admissible if

$$w \in W^{M_r}, \quad P\overline{N}_r g = P\overline{N}_r \overline{n}.$$

Let $f_{r,\overline{n},w} \in \mathrm{ind}_P^{P\overline{P}}(W)$ be the function supported on $P\overline{N}_r\overline{n}$ and equal to w on $\overline{N}_r\overline{n}$. The function $gf_{r,\overline{n},w} \in \mathrm{Ind}_P^G(W)$ is supported on $P\overline{N}_r\overline{n}g^{-1}$.

We fix an element $\Phi \in \mathrm{Hom}_{R[\overline{P}]}(\mathrm{ind}_P^{P\overline{P}} W, V)$. We show that Φ belongs to the image of J if W is z-locally finite following Emerton's method [Emerton, 4.4.6, resp. 4.4.3] in two steps:

(1) Φ belongs to the image of J when $\Phi(gf_{r,\overline{n},w}) = g\Phi(f_{r,\overline{n},w})$ for all admissible $(g, r, \overline{n}, w)$.
(2) $\Phi(gf_{r,\overline{n},w}) = g\Phi(f_{r,\overline{n},w})$ for all admissible $(g, r, \overline{n}, w)$ if W is z-locally finite.

Proof of (1) Let $g_1, \ldots, g_n$ in G and non-zero functions $f_1, \ldots, f_n$ in $\mathrm{ind}_P^{P\overline{P}}(W)$. We show that $\sum_i g_i\Phi(f_i) = \Phi(\sum_i g_i f_i)$. We choose $r \in \mathbb{N}$ large enough such that the f_i, viewed as elements of $C_c^\infty(\overline{N}, W)$, are left $\overline{N}_r$-invariant with values in W^{M_r}. We fix a subset X_r of G such that

$$G = \sqcup_{h \in X_r} P\overline{N}_r h, \quad P\overline{P} = \sqcup_{h \in X_r \cap \overline{N}} P\overline{N}_r h.$$

Let $Y_i \subset X_r \cap \overline{N}$ such that the support of f_i is $\sqcup_{\overline{n} \in Y_i} P\overline{N}_r\overline{n}$. For $n \in Y_i$, we have

$$f_i|_{P\overline{N}_r\overline{n}} = f_{r,\overline{n},f_i(\overline{n})}.$$

Since $G = \sqcup_{h \in X_r} P\overline{N}_r h g_i, f_i$ viewed as an element of $\mathrm{ind}_P^G W$ is equal to

$$f_i = \sum_{h \in X_r} f_i|_{P\overline{N}_r hg_i}$$

where $h \in X_r$ contributes to a nonzero term if and only if $P\overline{N}_r hg_i = P\overline{N}_r\overline{n}$ for some $\overline{n} \in Y_i$; when this happens $f_i|_{P\overline{N}_r hg_i} = f_{r,\overline{n},f_i(\overline{n})}$ hence $g_i\Phi(f_i|_{P\overline{N}_r hg_i}) = \Phi(g_i(f_i|_{P\overline{N}_r hg_i}))$ by the assumption of (1). We compute

$$\sum_i g_i\Phi(f_i) = \sum_h \sum_i g_i\Phi(f_i|_{P\overline{N}_r hg_i}) = \sum_h \sum_i \Phi(g_i(f_i|_{P\overline{N}_r hg_i}))$$

$$= \Phi\left(\sum_i g_i\left(\sum_h f_i|_{P\overline{N}_r h g_i}\right)\right) = \Phi\left(\sum_i g_i f_i\right).$$

Therefore $\sum_i g_i\Phi(f_i) = \Phi(\sum_i g_i f_i)$ for all $g_1,\dots,g_n$ in G and $f_1,\dots,f_n$ in $\mathrm{ind}_P^{P\overline{P}}(W)$, hence Φ belongs to the image of J.

Proof of (2). We assume $W \in \mathrm{Mod}_R^{z\text{-lf}}(M)$ and we prove $\Phi(gf_{r,\overline{n},w}) = g\Phi(f_{r,\overline{n},w})$. We reduce to $\overline{n} = 1$, as $f_{r,\overline{n},w} = \overline{n}^{-1}f_{r,1,w}$, $(g\overline{n}^{-1}, r, 1, w)$ is admissible, and Φ is $\overline{N}$-equivariant.

Let $(g, r, 1, w)$ admissible. We may suppose $w \neq 0$. We choose $(r', r'', a) \in \mathbb{Z}\times\mathbb{N}\times\mathbb{N}$ as follows. The integer $r' \in \mathbb{Z}$ depending on (g, r) is chosen so that the projection of the compact subset $\overline{N}_r g^{-1} \subset P\overline{N}_r$ onto N via the natural homeomorphism $P\overline{N} \to N \times \overline{P}$ is contained in $N_{r'}$, i.e., $\overline{N}_r g^{-1} \subset N_{r'}\overline{P}$. The integer $r'' \in \mathbb{N}$ depending on (r, w) and on our fixed element $z \in A_M$, is chosen so that the R-submodule of $V \in \mathrm{Mod}_R^\infty(G)$, generated by $\Phi(f_{r,1,w'})$ for w' in the finitely generated R-submodule $R[z]w$, is contained in $V^{K_{r''}}$, and $r'' \geq r$. Finally, the integer $a \in \mathbb{N}$ depending on r, r'' is chosen so that $z^a N_{r'} z^{-a} \subset N_{r''} \subset N_r$.

Let $\overline{v} \in \overline{N}_r$. The set $Pz^{-a}\overline{N}_r z^a\overline{v} = P\overline{N}_r z^a\overline{v}$ is contained in $P\overline{N}_r$ as $z^{-1} \in A_M$ contracts $\overline{N}$. The restriction of $f_{r,1,w}$ to $P\overline{N}_r z^a\overline{v}$ is $f_{r,z^a\overline{v},z^a(w)}$. We deduce

$$f_{r,1,w} = \sum_{\overline{v}\in z^{-a}\overline{N}_r z^a\backslash\overline{N}_r} (z^a\overline{v})^{-1} f_{r,1,z^a(w)}.$$

We are reduced to prove $\Phi(gv^{-1}z^{-a}f_{r,1,z^a(w)}) = g\Phi(v^{-1}z^{-a}f_{r,1,z^a(w)})$. As Φ is left $\overline{P}$-equivariant, $g\Phi(v^{-1}z^{-a}f_{r,1,z^a(w)}) = gv^{-1}z^{-a}\Phi(f_{r,1,z^a(w)})$. The set $g\overline{N}_r$ is contained in $\overline{P}N_{r'}$ and we may write $g\overline{v}^{-1}z^{-a} = \overline{p}n_{r'}z^{-a}$ with $n_{r'} \in N_{r'}, \overline{p} \in \overline{P}$. Using again that Φ is left $\overline{P}$-equivariant, we are reduced to prove

$$\Phi(n_{r'}z^{-a}f_{r,1,z^a(w)}) = n_{r'}z^{-a}\Phi(f_{r,1,z^a(w)}).$$

Applying z^a, we are reduced to prove

$$\Phi(z^a n_{r'}z^{-a}f_{r,1,z^a(w)}) = z^a n_{r'}z^{-a}\Phi(f_{r,1,z^a(w)}).$$

Let $w' \in R[z]w$ and $\overline{v} \in \overline{N}_r$. The function $f_{r,1,w'}$ viewed in $\mathrm{Ind}_P^G(W)$, of support $P\overline{N}_r$ and equal to $w' \in W^{M_r}$ on $\overline{N}_r$, is fixed by K_r. The element $\Phi(f_{r,1,w'}) \in V$ is fixed by $K_{r''}$. As $z^a N_{r'} z^{-a} \subset N_{r''} \subset N_r$, both elements $f_{r,1,z^a(w)}$ and $\Phi(f_{r,1,z^a(w)})$ are fixed by $z^a n_{r'} z^{-a}$, and the equality is obvious. □

7 The Hecke Description of $R_{\overline{P}}^{\overline{P}P} : \mathrm{Mod}_R^\infty(P) \to \mathrm{Mod}_R^\infty(M)$

We keep the notation of the preceding section. The submonoid $M^+ \subset M$ contracting the open compact subgroup N_0 of N is the set of $m \in M$ such that $mN_0m^{-1} \subset N_0$; it contains the open compact subgroup M_0 of M. The union $\cup_{a\in\mathbb{N}} z^{-a}M^+$ is equal to M.

The right adjoint of the restriction functor $\mathrm{Mod}_R(M) \to \mathrm{Mod}_R(M^+)$ is the induction functor

$$I_{M^+}^M : \mathrm{Mod}_R(M^+) \to \mathrm{Mod}_R(M)$$

sending $W \in \mathrm{Mod}_R(M^+)$ to the module $I_{M^+}^M(W)$ of R-linear maps $\psi : M \to W$ such that $\psi(mx) = m\psi(x)$ for all $m \in M^+, x \in M$, where M acts by right translations. The smoothification of $I_{M^+}^M$ is the smooth induction functor

$$\mathrm{Ind}_{M^+}^M : \mathrm{Mod}_R^\infty(M^+) \to \mathrm{Mod}_R^\infty(M).$$

Definition 7.1. Let $V \in \mathrm{Mod}_R^\infty(P)$. The monoid M^+ acts on V^{N_0} by the Hecke action $(m, v) \mapsto h_{N_0,m}(v)$,

$$h_{N_0,m}(v) = \sum_{n\in N_0/mN_0m^{-1}} nmv \quad (m \in M^+, v \in V^{N_0}). \tag{10}$$

The Hecke action of M^+ on V^{N_0} is smooth because it extends the natural action of M_0 on V^{N_0}.

Theorem 7.1. *The functor*

$$V \mapsto \mathrm{Ind}_{M^+}^M(V^{N_0}) : \mathrm{Mod}_R^\infty(P) \to \mathrm{Mod}_R^\infty(M) \tag{11}$$

is right adjoint to the functor $\mathrm{ind}_{\overline{P}}^{\overline{P}P}$.

The theorem says that the functors $\mathrm{Ind}_{M^+}^M(-^{N_0})$ and $R_{\overline{P}}^{\overline{P}P}$ are isomorphic. Their z-locally finite parts are also isomorphic. The Emerton's ordinary functor Ord_P is the A_M-locally finite part of the functor $\mathrm{Ind}_{M^+}^M(-^{N_0})$:

$$\mathrm{Ord}_P = (\mathrm{Ind}_{M^+}^M(-^{N_0}))^{A_M\text{-lf}} : \mathrm{Mod}_R^\infty(P) \to \mathrm{Mod}_R^{A_M\text{-lf}}(M),$$

or also the functor $\mathrm{Ord}_P^G := \mathrm{Ord}_P \circ \mathrm{Res}_P^G : \mathrm{Mod}_R^\infty(G) \to \mathrm{Mod}_R^{A_M\text{-lf}}(M)$. Applying Theorem 6.1, we obtain:

Corollary 7.3. *The functor* $R_{\overline{P}}^{G,z\text{-lf}}$ *is isomorphic to the functor*

$$V \mapsto (\mathrm{Ind}_{M^+}^{M}(V^{N_0}))^{z\text{-lf}} : \mathrm{Mod}_R^\infty(G) \to \mathrm{Mod}_R^{z\text{-lf}}(M)$$

The functor $R_{\overline{P}}^{G,A_M\text{-lf}}$ *is isomorphic to the Emerton's ordinary functor* Ord_P^G.

To prove that $(\mathrm{ind}_{\overline{P}}^{\overline{P}P}, \mathrm{Ind}_{M^+}^{M}(-^{N_0}))$ is an adjoint pair, we view $\mathrm{ind}_{\overline{P}}^{\overline{P}P}$ as

$$C_c^\infty(N,R) \otimes_R - : \mathrm{Mod}_R^\infty(M) \to \mathrm{Mod}_R^\infty(P),$$

where $P = MN$ acts on $C_c^\infty(N,R)$ by:

$$mf : x \mapsto f(m^{-1}xm),\ nf : x \mapsto f(xn), \quad (m,n,f) \in M \times N \times C_c^\infty(N,R).$$

(In particular $m1_{N_0} = 1_{mN_0m^{-1}}, n1_{N_0} = 1_{N_0n^{-1}}$). The right adjoint is well known:

Lemma 7.4. *The smoothification of the functor*

$$\mathrm{Hom}_{R[N]}(C_c^\infty(N,R), -) : \mathrm{Mod}_R^\infty(P) \to \mathrm{Mod}_R(M)$$

is the right adjoint of the functor $\mathrm{ind}_{\overline{P}}^{\overline{P}P}$.

The following proposition 7.5 implies that the functors $\mathrm{Hom}_{R[N]}(C_c^\infty(N,R), -)$ and

$$I_{M^+}^{M}(-^{N_0}) : \mathrm{Mod}_R^\infty(P) \to \mathrm{Mod}_R(M).$$

are isomorphic. Therefore the same is true for their smoothifications, $R_{\overline{P}}^{\overline{P}P}$ and $\mathrm{ind}_{M^+}^{M}(-^{N_0})$, and Theorem 7.1 is proved.

Let $V \in \mathrm{Mod}_R^\infty(P)$. We check that the value at 1_{N_0}

$$f \mapsto f(1_{N_0}) : \mathrm{Hom}_{R[N]}(C_c^\infty(N,R), V) \to V^{N_0}$$

is M^+-equivariant. As usual, $p \in P$ acts on f by $pf = p \circ f \circ p^{-1}$. In particular, for $m \in M$,

$$(mf)(1_{N_0}) = mf(m^{-1}1_{N_0}) = mf(1_{m^{-1}N_0m}).$$

For $m \in M^+$, we obtain

$$\begin{aligned}(mf)(1_{N_0}) &= m \sum_{n^{-1} \in N_0 \backslash m^{-1}N_0m} f(1_{N_0n^{-1}}) = \sum_{n^{-1} \in N_0 \backslash m^{-1}N_0m} mnf(1_{N_0}) \\ &= \sum_{n \in N_0/mN_0m^{-1}} nmf(1_{N_0}) = h_{N_0,m}(f(1_{N_0}))\ .\end{aligned}$$

By the adjunction $(\mathrm{Res}^M_{M^+}, I^M_{M^+})$, the value at 1_{N_0} induces an M-equivariant map

$$\Phi : \mathrm{Hom}_{R[N]}(C_c^\infty(N,R), V) \to I^M_{M^+}(V^{N_0}) \quad f \mapsto \Phi(f)(m) = (mf)(1_{N_0}) \ (m \in M). \tag{12}$$

Proposition 7.5. *The map Φ is an isomorphism of $R[M]$-modules.*

Proof. Φ is injective because the $R[P]$-module $C_c^\infty(N,R)$ is generated by 1_{N_0}. Indeed let $f \in \mathrm{Hom}_{R[N]}(C_c^\infty(N,R), V)$ such that $\Phi(f) = 0$. Then $f_\psi(m1_{N_0}) = f(1_{m^{-1}N_0 m}) = 0$ for all $m \in M$. As f is N-equivariant, $0 = f((mn)^{-1}1_{N_0}) = f(1_{m^{-1}N_0 mn})$ for all $n \in N$, hence $f = 0$.

Φ is surjective because for $\psi \in I^M_{M^+}(V^{N_0})$, there exists $f_\psi \in \mathrm{Hom}_{R[N]}(C_c^\infty(N,R), V)$ such that $f_\psi(m1_{N_0}) = m(\psi(m^{-1}))$ for all $m \in M$. We have $\Phi(f_\psi) = \psi$. The function f_ψ exists because, for all $a \in \mathbb{N}$,

$$z^a(\psi(z^{-a})) = z^a(\psi(zz^{-a-1})) = \sum_{n \in z^a N_0 z^{-a}/z^{a+1}N_0 z^{-a-1}} nz^{a+1}(\psi(z^{-a-1})).$$

(Note that the $R[N]$-module $C_c^\infty(N,R)$ is generated by $(1_{z^a N_0 z^{-a}})_{a\in\mathbb{N}}$, and that the values at $1_{z^a N_0 z^{-a}} = z^a 1_{N_0}$ identify $\mathrm{Hom}_{R[N]}(C_c^\infty(N,R), V)$ with the set of sequences $(v_a)_{a\in\mathbb{N}}$ in V such that $v_a = \sum_{n\in z^a N_0 z^{-a}/z^{a+1}N_0 z^{-a-1}} n v_{a+1}$.) □

Remark 7.6. For $V \in \mathrm{Mod}_R^\infty(P)$, a z^{-1}-finite element $\varphi \in I^M_{M^+}(V^{N_0})$ is smooth:

$$(\mathrm{Ind}^M_{M^+}(V^{N_0}))^{z^{-1}\text{-lf}} = (I^M_{M^+}(V^{N_0}))^{z^{-1}\text{-lf}}.$$

Proof. By hypothesis $R[z^{-1}]\varphi$ is contained in a finitely generated R-submodule W_φ of $I^M_{M^+}(V^{N_0})$. The image of W_φ by the map $f \mapsto f(1)$ is a finitely generated R-submodule of V^{N_0} containing $\varphi(z^{-a})$ for all $a \in \mathbb{N}$. Since the Hecke action of M^+ on V^{N_0} is smooth, there exists a large integer $r \in \mathbb{N}$ such that M_r fixes $\varphi(z^{-a})$ for all $a \in \mathbb{N}$. As $M = \cup_{a\in\mathbb{N}} M^+ z^{-a}$, two elements of $I^M_{M^+}(V^{N_0})$ equal on z^{-a} for all $a \in \mathbb{N}$ are equal. Hence φ is fixed by M_r, φ is smooth. □

Remark 7.7. Let $W \in \mathrm{Mod}_R^\infty(M^+)$ and $r \in \mathbb{N}$. An element $f \in I^M_{M^+}(W)$ is fixed by M_r if and only if $f(z^a)$ is fixed by M_r for all $a \in \mathbb{Z}$. The map

$$f \mapsto f|_{z^\mathbb{Z}} : (I^M_{M^+} W)^{M_r} \to I^{z^\mathbb{Z}}_{z^\mathbb{N}}(W^{M_r})$$

is an $R[z^\mathbb{Z}]$-isomorphism.

Proof. This is an easy consequence of $(m_r f)(m^+ z^a) = f(m^+ z^a m_r) = f(m^+ m_r z^a) = m^+ m_r(f(z^a))$ for $(m^+, m_r, a) \in M^+ \times M_r \times \mathbb{Z}$. □

8 The Right Adjoint $\mathrm{Ord}_{\overline{P}}$ of $\mathrm{Ind}_P^G : \mathrm{Mod}_R^{\mathrm{adm}}(M) \to \mathrm{Mod}_R^{\mathrm{adm}}(G)$

Theorem 8.1. *For $V \in \mathrm{Mod}_R^{\mathrm{adm}}(G)$, the representation $(I_{M+}^M(V^{N_0}))^{z^{-1}\text{-lf}}$ of M is admissible.*

Proof. By Remark 7.6, the representation $(I_{M+}^M(V^{N_0}))^{z^{-1}\text{-lf}}$ of M is smooth. Let $r \in \mathbb{N}$. Note that M_rN_0 is a group. By Remark 7.7, the map $f \mapsto f|_{z^{\mathbb{Z}}}$ is an $R[z^{\mathbb{Z}}]$-isomorphism from the M_r-fixed elements of $(I_{M+}^M(V^{N_0}))^{z^{-1}\text{-lf}}$ to

$$X = (I_{z^{\mathbb{N}}}^{z^{\mathbb{Z}}}(V^{N_0M_r}))^{z^{-1}\text{-lf}}.$$

We have $X \subset I_{z^{\mathbb{N}}}^{z^{\mathbb{Z}}}(Y)$ where Y is the image of X by $f \mapsto f(1)$, and is a $z^{\mathbb{N}}$-submodule of $V^{N_0M_r}$ (for the Hecke action) containing $f(z^a)$ for all $a \in \mathbb{Z}$. We have the compact open subgroup $\overline{N}_rM_rN_0$. We will prove (Proposition 8.2) that

$$Y \subset V^{\overline{N}_rM_rN_0}.$$

Admitting this, Y is a finitely generated R-module because V is admissible and R is noetherian. The action $h_{N_0,z}$ of z on Y is surjective because, for $f \in X$ we have $f(1) = f(zz^{-1}) = h_{N_0,z}f(z^{-1})$. A surjective endomorphism of a finitely generated R-module is bijective (this is an application of Nakayama lemma [Matsumura, Theorem 2.4]). Hence the action of z on Y is bijective. Hence $Y \simeq I_{z^{\mathbb{N}}}^{z^{\mathbb{Z}}}(Y)$ is a finitely generated R-module. As R is noetherian, X is a finitely generated R-module. Therefore $(I_{M+}^M(V^{N_0}))^{z^{-1}\text{-lf}}$ is admissible. □

Proposition 8.2. *If $f \in (I_{z^{\mathbb{N}}}^{z^{\mathbb{Z}}}(V^{M_rN_0}))^{z^{-1}\text{-lf}}$, then $f(1) \in V^{\overline{N}_rM_rN_0}$.*

Proof. We have

$$V^{M_rN_0} = \cup_{t \geq r} V^{\overline{N}_tM_rN_0}, \tag{13}$$

where $\overline{N}_tM_rN_0 = K_tM_rN_0 \subset G$ is a compact open subgroup as $M_rN_0 \subset K_0$ normalizes K_t, and the sequence $(\overline{N}_tM_rN_0)_{t \geq r}$ is strictly decreasing of intersection M_rN_0. We write $n(r,t) \in \mathbb{N}$ for the smallest integer such that $z^{-n}\overline{N}_rz^n \subset \overline{N}_t \subset \overline{N}_r$ for $n \geq n(r,t)$. The proof of the proposition is split into three steps.

(1) *$h_{N_0,z^n}(V^{\overline{N}_tM_rN_0})$ is fixed by $\overline{N}_rM_rN_0$ when $n \geq n(r,t)$.*

Let $v \in V^{\overline{N}_tM_rN_0}$ and $n \geq n(r,t)$. The element z^nv is fixed by $\overline{N}_rM_r$ as $\overline{N}_rM_rz^n \subset z^n\overline{N}_tM_r$. Let $\overline{n}_r \in \overline{N}_r$ and $(n_i)_{i\in I}$ a system of representatives of $N_0/z^nN_0z^{-n}$. Using the Iwahori decomposition $\overline{N}_rM_rN_0 = N_0\overline{N}_rM_r$ we write $\overline{n}_rn_i = n_i'\overline{b}_i$ with $n_i' \in N_0, \overline{b}_i \in \overline{N}_rM_r$. We compute:

$$\overline{n}_r h_{N_0,z^n}(v) = \sum_{i\in I} \overline{n}_r n_i z^n v = \sum_{i\in I} n'_i \overline{b}_i z^n v = \sum_{i\in I} n'_i z^n v. \tag{14}$$

We show that $(n'_i)_{i\in I}$ is a system of representatives of $N_0/z^n N_0 z^{-n}$, hence that $\overline{n}_r$ fixes $h_{N_0,z^n}(v)$, hence (1). We have to prove that $n_i'^{-1} n'_j \in z^n N_0 z^{-n}$ implies $i = j$. We write $n_i'^{-1} n'_j = \overline{b}_i n_i^{-1} n_j \overline{b}_j^{-1}$ and we assume that $\overline{b}_i n_i^{-1} n_j \overline{b}_j^{-1} \in z^n N_0 z^{-n}$. Then $n_i^{-1} n_j$ belongs to the group generated by $\overline{N}_r M_r$ and $z^n N_0 z^{-n}$, which is contained in the group $z^n \overline{N}_r M_r N_0 z^{-n}$. Hence $n_i^{-1} n_j \in z^n N_0 z^{-n}$. This implies $i = j$.

(2) $V^{\overline{N}_t M_r N_0}$ is stable by $h_{N_0,z}$ (hence by h_{N_0,z^n} for $n \in \mathbb{N}$).
When $t = r$, this follows from (1) because $n(t,t) = 0$. This is true for any large $t = r$. Hence the intersection $V^{M_r N_0} \cap V^{\overline{N}_t M_t N_0}$ is stable by $h_{N_0,z}$. But this intersection is $V^{\overline{N}_t M_r N_0}$ because the group generated by $M_r N_0$ and $\overline{N}_t M_t N_0$ is $\overline{N}_t M_r N_0$, as M_r contains M_t and normalizes $\overline{N}_t, M_t, N_0$. Hence (2).

(3) Let f be a z^{-1}-finite element of $I_{z^{\mathbb{N}}}^{z^{\mathbb{Z}}}(V^{M_r N_0})$. The R-module generated by $f(z^{-a})$ for $a \in \mathbb{N}$ is contained in a finitely generated R-submodule of $V^{M_r N_0}$. There exists $t \geq r$ such that $f(z^{-a})$ is contained in $V^{\overline{N}_t M_r N_0}$ for all $a \in \mathbb{N}$. By (2), $f \in I_{z^{\mathbb{N}}}^{z^{\mathbb{Z}}}(V^{\overline{N}_t M_r N_0})$. We have $f(1) \in \cap_{n\geq 1} h_{N_0,z^n}(V^{\overline{N}_t M_r N_0})$. By (1), $h_{N_0,z^n}(V^{\overline{N}_t M_r N_0}) \subset V^{\overline{N}_r M_r N_0}$ when $n \geq n(r,t)$. Hence $f(1) \in V^{\overline{N}_r M_r N_0}$. The proposition is proved.

□

This ends the proof of Theorem 8.1. An admissible representation of M is A_M-locally finite (Lemma 3.6). By Theorem 8.1, Remark 7.6, and Corollary 7.3, we deduce:

Corollary 8.3. *The (admissible) parabolic induction* $\mathrm{Ind}_P^G : \mathrm{Mod}_R^{\mathrm{adm}}(M) \to \mathrm{Mod}_R^{\mathrm{adm}}(G)$ *admits a right adjoint, equal to*

$$(R_P^G)^{A_M\text{-lf}} \simeq \mathrm{Ord}_{\overline{P}}^G : \mathrm{Mod}_R^{\mathrm{adm}}(G) \to \mathrm{Mod}_R^{\mathrm{adm}}(M).$$

Corollary 8.4. *When p is nilpotent in R, the admissible parabolic induction* Ind_P^G *is fully faithful, and the unit* $\mathrm{id} \mapsto \mathrm{Ord}_{\overline{P}}^G \circ \mathrm{Ind}_P^G$ *of the adjunction* $(\mathrm{Ind}_P^G, \mathrm{Ord}_{\overline{P}}^G)$ *is an isomorphism.*

Proof. Lemma 4.6, Corollary 5.3. □

It is not known if the N-coinvariant functor respects admissibility when the characteristic of F is p. When R is a field where p is invertible, the N-coinvariant functor respects admissibility. For the convenience of the reader, we give the proof which is a variant of the proof of [Viglivre, II.3.4].

(i) Let R be a commutative ring (we do not assume that R is noetherian) and $V \in \mathrm{Mod}_R^\infty(G)$. For $v \in V^{N_0}$ and $a \in \mathbb{N}$, we have $h_{N_0,z^a}(v) = \sum_{n\in N_0/z^a N_0 z^{-a}} n z^a v = z^a \sum_{n\in z^{-a} N_0 z^a/N_0} nv$. Applying the map $\kappa : V \to V_N$, we get

$$\kappa(h_{N_0,z^a}(v)) = [N_0 : z^a N_0 z^{-a}] z^a \kappa(v). \tag{15}$$

The index $[N_0 : z^a N_0 z^{-a}]$ is a power of p which goes to infinity with a. (Note that when a power of p vanishes in R, $\kappa(h_{N_0,z^a}(v)) = 0$ when a is large.) For $r \in \mathbb{N}$ we have $\kappa(V^{M_r N_0}) \subset (V_N)^{M_r}$ because $m\kappa(v) = \kappa(mv)$ for $m \in M, v \in V$.

(ii) We assume now that p is invertible in R. The above inclusion for $r \in \mathbb{N}$ is an equality

$$\kappa(V^{M_r N_0}) = (V_N)^{M_r}. \tag{16}$$

Indeed, let $w \in (V_N)^{M_r}$ and $v \in V$ with $\kappa(v) = w$. The fixator H_r of v in the pro-p-group $M_r N_0$ is open of index a power of p. The element $[M_r N_0 : H_r]^{-1} \sum_{b \in M_r N_0 / H_r} bv$ is well defined, is fixed by $M_r N_0$, and has image w in V_N. Hence (16). As V_N is a smooth representation of M and $V^{N_0} = \cup_{r \in \mathbb{N}} V^{M_r N_0}$, (16) implies $\kappa(V^{N_0}) = V_N$ and by (13),

$$\cup_{t \geq r}\, \kappa(V^{\overline{N}_t M_r N_0}) = (V_N)^{M_r}. \tag{17}$$

Assume $a \geq n(r,t)$, by (15) and by the proof of Proposition 8.2,

$$z^a \kappa(V^{\overline{N}_t M_r N_0}) = \kappa(h_{N_0,z^a}(V^{\overline{N}_t M_r N_0})) \subset \kappa(V^{\overline{N}_r M_r N_0}). \tag{18}$$

If X is a finitely generated R-submodule of $V_N^{M_r}$, there exists $t \in \mathbb{N}$ such that $X \subset \kappa(V^{\overline{N}_t M_r N_0})$, hence by (18) there exists $a \in \mathbb{N}$ such that

$$z^a X \subset \kappa(V^{\overline{N}_r M_r N_0}). \tag{19}$$

(iii) We assume now that R is a field where p is invertible and $V \in \mathrm{Mod}_R^{\mathrm{adm}}(G)$. By (19) the dimensions of the finite dimensional subspaces of $V_N^{M_r}$ are bounded, hence $V_N^{M_r}$ is finite dimensional. This is true for all $r \in \mathbb{N}$ therefore $V_N \in \mathrm{Mod}_R^{\mathrm{adm}}(M)$.

References

[Bu] C.J. Bushnell, Representations of reductive p-adic groups: localization of Hecke algebras and applications. J. Lond. Math. Soc. **63**(2), 364–386 (2001)

[Bo] A. Borel, *Linear Algebraic Groups*, 2 edn. (Springer, New York, 1991).

[BT1] F. Bruhat, J. Tits, Groupes réductifs sur un corps local: I. Données radicielles valuées. Publ. Math. I.H.E.S. **41** (1972)

[Dat] J.-F. Dat, Finitude pour les représentations lisses de groupes réductifs p-adiques. J. Inst. Math. Jussieu **8**(1), 261–333 (2009)

[Emerton] M. Emerton, Ordinary parts of admissible representations of p-adic reductive groups I. Astérisque **331**, 355–402 (2010)

[Emerton2] M. Emerton, Ordinary parts of admissible representations of p-adic reductive groups II. Astérisque **331**, 403–459 (2010)

[HV] G. Henniart, M.-F. Vigneras, Comparison of compact induction with parabolic induction. Special issue to the memory of J. Rogawski. Pac. J. Math. **260**(2), 457–495 (2012)

[KS] M. Kashiwara, P. Schapira, Categories and sheaves. *Grundlehren des mathematischen Wissenschaften*, vol. 332 (Springer, New York, 2006). Errata in http://webusers.imj-prg.fr/~pierre.schapira,/books.

[Ly] T. Ly, Irreducible representations modulo p representations of $GL(n, D)$. Thesis. University of Paris 7, 2013

[Matsumura] H. Matsumura, *Categories for the Working Mathematician* (Springer, New York, 1971)

[ML] L.S. Mac, *Categories for the Working Mathematician* (Springer, New York, 1971)

[SVZ] P. Schneider, M.-F. Vigneras, G. Zabradi, From étale P_+-representations to G-equivariant sheaves on G/P (2012). Preprint

[SGA4] A. Grothendieck, J.-L. Verdier, Préfaisceaux *Séminaire de géometrie algébrique by M. Artin, A. Grothendieck and J.-L. Verdier (1963–1964)*. Lecture Notes in Mathematics, vol. 269 (Springer, New York, 1972)

[VigLang] M.-F. Vigneras, Représentations p-adiques de torsion admissibles. *Number Theory, Analysis and Geometry: In Memory of Serge Lang* (Springer, New York, 2011)

[Viglivre] M.-F. Vigneras, Représentations l-modulaires d'un groupe réductif p-adique avec $\ell \neq p$. *Progress in Mathematics*, vol. 131 (Birkhäuser, Boston, 1996)