CLEANLINESS OF GEODESICS IN HYPERBOLIC 3-MANIFOLDS

C. Adams, A. Colestock, J. Fowler, W. Gillam, and E. Katerman

In this paper, we investigate geodesics in cusped hyperbolic 3-manifolds. We derive conditions guaranteeing the existence of geodesics avoiding the cusps and use these geodesics to show that in “almost all” finite volume hyperbolic 3-manifolds, infinitely many horoballs in the universal cover corresponding to a cusp are visible in a fundamental domain of the cusp when viewed from infinity.

1. Introduction.

Let $M$ be a cusped hyperbolic manifold of finite volume. Lift $M$ to the upper-half space model of $\mathbb{H}^3$, so that cusps lift to horoballs. Center one of the horoballs at infinity, then expand one cusp until it just touches itself or another cusp. Inflate each cusp in turn until all have maximal size without intersecting themselves or another cusp. The result is a maximal cusp or maximal set of cusps. Looking down from infinity, we see a pattern of horoballs within the fundamental parallelogram for the cusp subgroup that fixes $\infty$. Call the set of horoballs in the parallelogram the horoball diagram for the manifold. If the horoballs are opaque, how many are visible? Do finitely many horoballs suffice to cover the fundamental parallelogram? This question, first asked by Darren Long, is answered below.

There are cusped hyperbolic 3-manifolds in which a viewer situated at infinity can see only finitely many horoballs in the horoball diagram. This is true of the figure-eight knot complement because the first three layers of horoballs cover the plane (see Figure 1). Since this is not obvious from a diagram, we prove this fact in Section 4. And, because any singly cusped finite cover of the figure-eight knot complement will also have this property, there are infinitely many examples of manifolds with this property. But remarkably, such examples are exceptional:

**Theorem 1.1.** In the horoball diagram of almost all finite volume hyperbolic 3-manifolds, infinitely many horoballs are visible from infinity.
In Section 2, we define “almost all” and discuss the notation used throughout the paper. In Section 3, we prove Theorem 1.1. In Section 4, we investigate the properties of the figure-eight knot complement and the Whitehead link complement. Finally, we explain various implications of these results in Section 5 by relating our ideas to number theoretic problems. We are grateful to Ian Agol for bringing these questions to our attention.

2. Preliminaries.

Here we list the conventions used in this paper. We will be working in the upper-half plane model of $\mathbb{H}^2$ and the upper-half space model of $\mathbb{H}^3$, and points will be denoted $p = (z, t) \in \mathbb{H}^3$ (or $p = (x, t) \in \mathbb{H}^2$) where $x, t \in \mathbb{R}, z \in \mathbb{C}$ and $t > 0$; that is, $\partial \mathbb{H}^3 = \mathbb{C} \cup \{\infty\}$ and $\partial \mathbb{H}^2 = \mathbb{R} \cup \{\infty\}$. Given two points $p_1, p_2 \in \mathbb{H}^n$, the Euclidean length between $p_1$ and $p_2$ will be denoted $\ell_e(p_1, p_2)$. The hyperbolic length between points $p_1$ and $p_2$ will be denoted $\ell(p_1, p_2)$. A cusp in a hyperbolic 3-manifold is a subset with interior homeomorphic to $T^2 \times (0, 1)$ which lifts to a set of horoballs in $\mathbb{H}^3$. Choosing one such horoball to be centered at $\infty$, the cusp subgroup is the $\mathbb{Z} + \mathbb{Z}$ subgroup of the fundamental group of the manifold realized as a discrete group of fixed point free isometries. A fundamental domain for the action of the cusp subgroup on the boundary plane $\mathbb{C}$ is a parallelogram $P$.

Horocycles or horodiscs in $\mathbb{H}^2$ will be denoted by $h$ or $h_i$, and horospheres or horoballs in $\mathbb{H}^3$ by $H$ and $H_i$. In both cases, the corresponding Euclidean radii will be denoted by $r, r_i, r, r_i \in \mathbb{R}^+$ respectively.

Definition 2.1. We say that almost all finite volume hyperbolic 3-manifolds have a property $Q$ if, for any real number $V > 0$, all but finitely many with volume less than $V$ have property $Q$.

Stated differently, “almost all” is “all but a finite number below any given volume.”
Definition 2.2. A geodesic is **singly orthogonal** if it intersects the cusp boundary exactly once and does so at a right angle; a **doubly orthogonal** geodesic intersects the cusp boundary twice perpendicularly.

That is, a singly orthogonal geodesic is perpendicular to the tangent plane of the cusp at the point of intersection. Note that a geodesic entering the cusp perpendicularly stays in the cusp from then on.

A cusped hyperbolic 3-manifold has infinitely many doubly orthogonal geodesics connecting any cusp to itself and infinitely many connecting any cusp to any other cusp. We can see this by lifting cusps to horoballs, centering a horoball at infinity, and then we find infinitely many doubly orthogonal geodesics by drawing the vertical geodesics connecting the centers of horoballs and infinity. If we are allowed to alter the size of the cusps, we shall choose a configuration such that no cusp overlaps the interior of itself or another cusp. Below, we will use $\ell_h$ to denote the hyperbolic length of a doubly orthogonal geodesic outside the cusp or cusps. Usually, the size of the cusps is chosen to maximize cusp volume, and in the case of a 1-cusped hyperbolic 3-manifold, this choice is canonical. Given any geodesic in a cusped hyperbolic 3-manifold, we have the following definition:

**Definition 2.3.** A **dirty** geodesic intersects the cusp boundary non-orthogonally at least once, a **spotted** geodesic intersects the cusp boundary tangentially at least once and perpendicularly otherwise, and a **clean** geodesic is neither dirty nor spotted.

![Figure 2.](image)

Figure 2. The various possibilities for a geodesic intersecting the cusp.

For examples of lifts of geodesics of each type see Figure 2. We will find that the existence of geodesics of these various types is intimately related to the question of horoball visibility.
3. Horoball visibility and cleanliness.

We want to show that for almost all hyperbolic 3-manifolds, infinitely many horoballs are visible. The proof consists of three parts. First, we show that infinitely many horoballs are visible if there is a clean singly orthogonal geodesic. Second, there is a clean singly orthogonal geodesic if there is a "short" closed geodesics. Third, there are "short" closed geodesics in almost all hyperbolic 3-manifolds.

**Theorem 3.1.** If a hyperbolic 3-manifold $M$ contains a clean singly orthogonal geodesic then infinitely many horoballs in the horoball diagram are visible from $\infty$.

**Proof.** Suppose that $M$ contains a clean singly orthogonal geodesic $\alpha$. Let $P$ be a parallelogram in $\mathbb{C}$ that is a fundamental domain for the cusp subgroup and let $\tilde{\alpha}$ be a lift of $\alpha$ to $\mathbb{H}^3$ with one endpoint at infinity and the other endpoint in $P$. Suppose that only finitely many horoballs are visible. Then there is some $\epsilon > 0$ such that all horoballs $H^*_i$ with radius $r^*_i < \epsilon$ are in the shadows of the horoballs $H_i$ of radius $r_i \geq \epsilon$. Then there is a neighborhood of $\tilde{\alpha}$ of Euclidean radius $\delta > 0$ which is disjoint from all the $H_i$. Since horoball centers are dense in $\mathbb{C}$, there must be a horoball $H$ whose center lies in this neighborhood, and consequently $H$ has radius $r < \epsilon$. Since part of $H$ lies in the $\delta$-neighborhood of $\tilde{\alpha}$, the projection of $H$ to $\mathbb{C}$ is not completely covered by the projections of the $H_i$. 

Although the ability to see infinitely many horoballs from infinity may not imply the existence of a clean singly orthogonal geodesic, there is a partial converse to the above result.

**Theorem 3.2.** Given a hyperbolic 3-manifold $M$, if infinitely many horoballs in the horoball diagram are visible from $\infty$, then $M$ contains a clean or spotted singly orthogonal geodesic.

**Proof.** The fundamental parallelogram $P$ for the cusp subgroup of $M$ is a compact subset of $\mathbb{C}$. Let $H_0$ denote one of the largest horoballs with center in $P$. Choose $H_{i+1}$ to be one of the largest horoballs with center in $P$ which is smaller than $H_i$ and whose projection to $\mathbb{C}$ is not completely covered by the projections of larger horoballs. $H_i$ exists for all $i$ by assumption. Let $c_i$ denote the center of $H_i$; since $P$ is compact, we can pick $a \in P$ a limit point of $\{c_i\}$. Then $a$ cannot lie in the interior of the projection of a horoball $H$ because if it did, then we could choose an $N$ large enough so that $H_N$ is also completely contained within the projection of $H$. So, either $a$ is outside the projections of all horoballs or $a$ lies on the boundary of the projection(s) of one or more horoballs. We conclude that the geodesic from $a$ to $\infty$ covers a clean singly orthogonal or spotted singly orthogonal geodesic in the manifold. 

$\square$
Corollary 3.3. If a hyperbolic 3-manifold $M$ contains infinitely many clean doubly orthogonal geodesics then $M$ contains a clean or spotted singly orthogonal geodesic.

Proof. If $M$ contains infinitely many clean doubly orthogonal geodesics, then the fundamental parallelogram for $M$ in $\mathbb{H}^3$ contains infinitely many horoballs such that the geodesic from the top of each horoball to the horoball at infinity does not touch any other horoballs. Since the projection of each such horoball to $\mathbb{C}$ contains a region uncovered by the other horoballs, the result follows immediately from Theorem 3.2.

There are many relationships between geodesic length and clean geodesics. For example, any doubly orthogonal geodesic $g$ of hyperbolic length $\ell_h(g) < \log 4$ is clean. Some geometry shows that there is not enough room for a third horoball to fit between the two horoballs pierced by the geodesic. (see Figure 3).

![Figure 3](image.png)

Figure 3. If a doubly orthogonal geodesic is short enough, then no horoball can intersect it.

We also look at simple closed geodesics, and define them to be clean if they avoid the cusps.

Lemma 3.4. Any closed geodesic $g$ of hyperbolic length $\ell_h(g) < \log(2+2\sqrt{2})$ is clean.

Proof. We must prove that if $g$ is short enough, then there is a tubular neighborhood around $g$ that misses all horoballs. Let $g$ be the closed geodesic lifting to a vertical geodesic $\tilde{g}$, so that $g$ has endpoints $(0, 1)$ and $(0, 1 + \ell_c g)$ and $\tilde{g}$ has ideal endpoints $(0, 0)$ and $\infty$. Let $H_1$ denote the horoball centered at one endpoint of the semi-circular geodesic which intersects $\tilde{g}$ at $(0, 1)$, and let $H_2$ be the horoball centered a Euclidean distance $\ell_c(g) + 1$ from the origin on the same line as the origin and the center of $H_1$. We want to find $g$ so that when we inflate both $H_1$ and $H_2$, then the moment they become tangent to each other will be the exact moment when they each become tangent to $g^*$. We know that $r_1 = 1$, $r_2 = \ell_c(g) + 1$, and the distance between the centers of $H_1$ and $H_2$ is $d(H_1, H_2) = \ell_c(g)$. Thus, from the triangle whose
hypoteneuse is the line between the Euclidean centers of $H_1$ and $H_2$, we find that $\ell_e(g) = 2 + 2\sqrt{2}$.

Surprisingly, if a manifold has a sufficiently short closed geodesic, there is a clean singly orthogonal geodesic. Before we prove the theorem for 3-manifolds, we will first prove the corresponding result for 2-manifolds.

**Lemma 3.5.** A cusped hyperbolic 2-manifold with a closed geodesic $g$ of hyperbolic length $\ell_h(g) < \log 3$ will have a clean singly orthogonal geodesic.

**Proof.** Lift the hyperbolic 2-manifold to the upper half-plane model, lifting $g$ to a vertical geodesic $\tilde{g}$. Expand a tubular neighborhood about $\tilde{g}$ until it just touches the cusp, say at a horocycle $h$. Draw the vertical geodesic $\tilde{g}'$ heading up from the center of $h$ and leaving the cusp orthogonally. Intuitively, if the tubular neighborhood is wide enough (i.e., the geodesic is short enough), then there will not be enough room above $h$ and below the tubular neighborhood for another horocycle to hit $\tilde{g}'$, which will make $\tilde{g}'$ a clean singly orthogonal geodesic. See Figure 4.

![Figure 4. A short geodesic with a wide neighborhood.](image)

Suppose $h$ has a Euclidean radius of $r$, and let $h'$ be the image of $h$ under the isometry generating the closed geodesic $g$, so $h'$ is a horocycle of radius $(1 + \ell_e(g))r$. For convenience, choose coordinates so that the vertical line $g$ sits at 0, and the horocycle $h$ has center at 1, so $h'$ has center at $(1 + \ell_e(g))$. Now $r$ is as large as possible when $h$ and $h'$ are tangent, which gives

$$r \leq \sqrt{\frac{\ell_e(g)^2}{4 + 4\ell_e(g)}}.$$
But if
\begin{equation}
(3.2) \quad r(1 + \ell_e(g)) \leq \ell_e(g),
\end{equation}
then $h'$ doesn’t hit $\tilde{g}'$. And if $h'$ doesn’t hit $\tilde{g}'$, then no horocycle can hit $\tilde{g}'$, because $h'$ was chosen to be maximal. Putting inequalities (3.1) and (3.2) together gives us the sufficient condition that $\ell_e(g) < 3$ (or equivalently that $r < 3/4$).

The proof for the 3-manifold case is similar, except that we have to take into account the loxodromic nature of a closed geodesic in $\mathbb{H}^3$. We utilize ideas similar to those appearing in [Mey86].

**Theorem 3.6.** Any cusped hyperbolic 3-manifold with a closed geodesic of length less than 0.1777 will have a clean singly orthogonal geodesic.

**Proof.** Lift the manifold to the upper half-space model in such a way that the given closed geodesic lifts to a vertical geodesic above the origin. Expand a tubular neighborhood of this closed geodesic until it just hits the cusp. Let $H$ be a horoball tangent to the neighborhood, and suppose the center of $H$ is at $(1, 0)$ and has radius $r$. Now let $\tilde{g}'$ be the vertical geodesic between $(0, 1)$ and $\infty$; we claim that $\tilde{g}'$ is the desired singly orthogonal clean geodesic.

Suppose a horoball hits $\tilde{g}'$. We rotate this horoball around $\tilde{g}'$ until the center of the horoball is at some positive real number. This situation is equivalent to the $\mathbb{H}^2$ case, and by the previous lemma, we know that another horoball cannot hit $\tilde{g}'$ provided $r < 3/4$.

We therefore want to bound $r$ based on $g$. We assume that $\theta$, the twist of $g$, is less than $2\pi/6$. Let $H'$ be the image of $H$ under the isometry corresponding to the closed geodesic $g$. The radius $r$ is maximized when $H$ and $H'$ are tangent. By the law of cosines, we bound the radius:
\[
r \leq \sqrt{\frac{1 + (1 + \ell_e(g))^2 - 2(1 + \ell_e(g))\cos \theta}{4 + 4\ell_e(g)}}.
\]
Assuming $-2\pi/6 < \theta < 2\pi/6$, we have $r < 3/4$ whenever $\ell_e(g) < (13 + \sqrt{105})/8$, which we will denote by $k$. Thus, if the hyperbolic length of $g$ is less than $\log k$ and the twist of $g$ is less than $2\pi/6$, then we have our singly orthogonal clean geodesic.

But by unwinding $g$ up to six times, we can always get a closed geodesic with twist less than $2\pi/6$. So if we have a closed geodesic with any twist and of hyperbolic length less than $(\log k)/6 \approx 0.1777$, then we will have a singly orthogonal clean geodesic.

It’s important to realize that our assumption of $\theta < 2\pi/6$ was not arbitrary; choosing $2\pi/6$ produces the best possible bound with this method.

**Lemma 3.7.** Given a real number $\epsilon > 0$, almost all hyperbolic 3-manifolds have a simple closed geodesic of length less than $\epsilon$. 

Proof. Pick a value $V > 0$. By the results of Thurston and Jørgensen (cf. [Thu79]), every hyperbolic 3-manifold of volume less than $V$ comes from surgery on a finite set of cusped hyperbolic 3-manifolds. Therefore, for any given bound $B$, all but finitely many of the manifolds of volume less than $V$ will have a pair of surgery coefficients $(p, q)$ that satisfy $p^2 + q^2 > B$. By choosing $B$ large, we can insure that the core of the surgery becomes a simple closed geodesic of length less than $\epsilon$. \hfill \Box

The following theorem completes the proof of the main result of this paper:

**Theorem 3.8.** In almost all hyperbolic 3-manifolds, infinitely many horoballs are visible looking down from infinity to the fundamental parallelogram.

**Proof.** By the previous lemma, almost all hyperbolic 3-manifolds have a geodesic $g$ where $\ell_h(g) < 0.17$. But by Theorem 3.6, a hyperbolic 3-manifold with a closed geodesic shorter than 0.17 has a clean singly orthogonal geodesic. And by Theorem 3.1, a clean singly orthogonal geodesic means that infinitely many horoballs are visible. \hfill \Box

4. Examples.

An example of a manifold that does not have a clean singly orthogonal geodesic is the figure-eight knot complement.

**Lemma 4.1.** The first three layers of horoballs that make up the horoball packing of the figure-eight knot complement project to cover $\mathbb{C}$ entirely.

**Proof.** Let $H_i$ denote the $i^{th}$ largest horoball in the cusp diagram of $K$ for $1 \leq i \leq 3$. Let $r_i$ be the radius of $H_i$ and let $d(H_i, H_j)$ be the distance between the centers of tangent horoballs $H_i$ and $H_j$ where the distance between the centers of $H_i$ and its nearest copy is computed if $i = j$. Since certain copies of $H_1, H_2,$ and $H_3$ are pairwise tangent, we can find $d(H_i, H_j)$ for $1 \leq i \leq 3, 1 \leq j \leq 3$. Since $H_1$ and its nearest translate are tangent to one another and to the horoball centered at infinity, we see that $d(H_1, H_1) = 1$. Next, we can find $d(H_1, H_2)$ by noting that the hexagonal packing of the copies of $H_1$ force the centers of the copies of $H_2$ to be in the center of the equilateral triangle with vertices at the centers of the copies of $H_1$. Since $r_1 = 1/2$, we find that

$$d(H_1, H_2) = \frac{1}{2} \frac{1}{\sin \frac{\pi}{6}} = \frac{1}{\sqrt{3}}.$$  

It is not hard to see that $(1/2 + r_2)^2 = (1/2 - r_2)^2 + 1/3$ and therefore that $r_2 = 1/6$. Similarly, since $d(H_1, H_3) = 1/2$, we find that $r_3 = 1/8$. 
Let $D_i$ denote the projection of the interior of $H_i$. Now, to show that the copies of $D_1, D_2, \text{ and } D_3$ completely cover $C$, it is sufficient to show that
\[ \bigcap_{i=1}^{3} D_i \neq \emptyset. \]
And, to show this, it suffices to show that $\theta_1 + \theta_2 > \pi/6$, where $\theta_1$ is the angle between the line segment joining the centers of $H_1$ and $H_2$ and the line segment joining the center of $H_1$ with the point where $\partial D_1$ and $\partial D_2$ intersect, and $\theta_2$ is similarly defined for $H_1$ and $H_3$. Now we have
\[ \frac{1}{36} = \frac{1}{3} + \frac{1}{4} - 2 \left( \frac{1}{\sqrt{3}} \right) \left( \frac{1}{2} \right) \cos \theta_1. \]
And thus, $\theta_1 = \cos^{-1}(\frac{5\sqrt{3}}{9})$, and similarly $\theta_2 = \cos^{-1}(\frac{31}{32})$. (For helpful pictures, see Figure 1 or refer to SnapPea [Wee].)  

**Corollary 4.2.** In the figure-eight knot complement:

1. There are no clean singly orthogonal geodesics.
2. There are only finitely many doubly orthogonal clean geodesics.

**Proof.** The first part follows immediately from the above theorem. An infinitely long clean geodesic can be pictured as a vertical geodesic leaving the horoball centered at $\infty$, but all vertical geodesics in the figure-eight knot complement must intersect the cusp a second time by the theorem.

The second part also follows immediately since there are only finitely many copies of the largest three horoballs in the fundamental domain for the cusp of the figure-eight knot complement in the universal cover. □

An example of a manifold that has a horoball pattern which does not completely cover $C$ and thus from which infinitely many horoballs are visible from infinity is the Whitehead link complement.

**Lemma 4.3.** The Whitehead link complement considered with symmetric tangent cusps has a clean singly orthogonal geodesic.

**Proof.** Take the set of horoballs in the upper-half-space model with centers at $p/q$ (reduced) where $p, q \in \mathbb{Q}(\sqrt{-1})$ and diameters given by $\frac{1}{k\gamma}$. The horoball centered at $\{\infty\}$ is represented by a horizontal plane at height $\frac{k}{2}$. This set of horoballs is invariant under the Picard group and therefore projects to the two cusps in the Whitehead link complement. Ford [For25] demonstrated that in such a space there exists a semi-circle $C$ with endpoints $(\frac{1}{2}, -\frac{\sqrt{3}}{2}, 0)$ and $(\frac{1}{2}, \frac{\sqrt{3}}{2}, 0)$ and radius $\frac{\sqrt{3}}{2}$, that is tangent to an infinite number of horoballs without entering the interior of any when $k = \sqrt{3}$. Although these horoballs are not disjoint, we can shrink back the resultant cusps in the Whitehead link at a symmetric rate until the horoball centered at infinity has boundary equal to the horizontal plane at height 1, at which point
all of the horoballs have disjoint interiors. Then there is a tube, \( t \), of radius \( \log \frac{2}{3} \) around \( C \) that does not intersect any horoballs in its interior. Let \( v \) be a singly orthogonal geodesic with basepoint \( \left( \frac{1}{2}, \frac{\sqrt{3}}{2} \right) \). Moving upwards towards \( \infty \), \( v \) travels within \( t \) until the point \( s = \left( \frac{1}{2}, \frac{\sqrt{3}}{2}, \frac{1}{4} \right) \) and thus must be clean within that region. Above \( s \), the only balls that might intersect \( v \) are those with diameter at least \( \frac{1}{4} \). But by looking at the balls centered at \( \frac{1+i}{2+i}, \frac{0+i}{1+2i}, \frac{1+i}{2+i}, \frac{1+0i}{2i+0i} \), we see that they are mutually tangent and enclose an area around \( v \) which does not cover its basepoint and prevents any balls of diameter at least \( \frac{1}{4} \) from being in that region. Thus \( v \) is a clean singly orthogonal geodesic. \( \square \)

\textbf{Theorem 4.4.} Infinitely many horoballs are visible from infinity within the fundamental parallelogram of the Whitehead link complement when considered with symmetric tangent cusps.

\textit{Proof.} This follows immediately from the above lemma and Theorem 3.1. \( \square \)

5. Number theory.

Many of these questions have relevance to number theory. In every cusped hyperbolic 3-manifold, the centers of the horoballs form a subfield of the complex plane. For example, if \( \Gamma \) is a torsion-free subgroup of finite index in a Bianchi group, then \( \mathbb{H}^3/\Gamma \) is a finite volume hyperbolic 3-manifold and questions about clean singly orthogonal geodesics are closely related to standard diophantine approximation.

Let \( \alpha \) be a real irrational number. In 1891, Hurwitz showed that the inequality

\[ |\alpha - \frac{p}{q}| < \frac{1}{k|q|^2} \]

has infinitely many solutions in coprime integers \( p \) and \( q \) when \( k = \sqrt{5} \), and that \( \sqrt{5} \) is the best constant possible. This inequality can be generalized to approximate complex numbers.

Let \( d \) be a positive square-free integer and let \( O_d \) be the ring of integers in \( \mathbb{Q}(\sqrt{-d}) \). Let \( \alpha \in \mathbb{C} - \mathbb{Q}(\sqrt{-d}) \). Denote by \( k_d(\alpha) \) the supremum of all \( k \) such that the inequality above has infinitely many solutions in \( p, q \in O_d \). Define the set of numbers

\[ L_d = \{ 1/k_d(\alpha) : \alpha \in \mathbb{C} - \mathbb{Q}(\sqrt{-d}) \} \]

to be the Lagrange spectrum for the imaginary quadratic number field \( \mathbb{Q}(\sqrt{-d}) \) and \( C_d = \sup L_d \) the Hurwitz constant for the field. See [For25] and [Vul].

Geometrically, this inequality indicates that if we expand the horoballs for a given manifold, \( M = \mathbb{H}^3/\Gamma \) where \( \Gamma \) is a torsion-free subgroup of finite
CLEANLINESS OF GEODESICS IN HYPERBOLIC 3-MANIFOLDS

index in a Bianchi group, until they all have radii equal to \( C_d |q|^2 \) then any vertical geodesic with basepoint \( \alpha \) will intersect an infinite number of horoballs. But horoballs with such radii are not necessarily disjoint. An interesting question which presents itself is, what happens if we allow the radii of the horoballs to correspond to those of a maximal cusp, thereby ensuring that the horoballs are disjoint? This is equivalent to studying the inequality when \( k = 1 \).

As we have shown, the plane is completely covered by the horoballs in a maximal cusp of the figure eight knot complement. Therefore, there exists at least one solution to the inequality for all \( \alpha \)'s when \( k = 1 \). With the Whitehead link complement, however, the existence of a clean singly orthogonal geodesic implies that if \( \alpha = \frac{1}{2} + \frac{\sqrt{3}}{2} i \) then the inequality has no solutions. These questions could be applied to other hyperbolic 3-manifolds besides those associated with the Bianchi groups.

References


Received October 1, 2002 and revised January 20, 2003. This research was supported by Williams College and the National Science Foundation under grants DMS-9820570 and DMS-9803362.

Dept. of Mathematics
Williams College
Williamstown, MA 01267
E-mail address: Colin.Adams@williams.edu

DEPT. OF MATHEMATICS
WILLIAMS COLLEGE
WILLIAMSTOWN, MA 01267

DEPT. OF MATHEMATICS
WILLIAMS COLLEGE
WILLIAMSTOWN, MA 01267

DEPT. OF MATHEMATICS
WILLIAMS COLLEGE
WILLIAMSTOWN, MA 01267

DEPT. OF MATHEMATICS
WILLIAMS COLLEGE
WILLIAMSTOWN, MA 01267
RINGS WITH SOVABLE QUIVERS

FRANK W. ANDERSON AND BARBARA K. D’AMBROSIA

An artinian ring $R$ is square-free in case none of its indecomposable projective modules has a repeated composition factor. Let $Q$ be the quiver of such a square-free ring $R$. In this paper we show that if $R$ is indecomposable and transitive on the cyclic components of $Q$ and if $Q$ contains no $n$-crown, then $R \cong D \otimes_K A$ where $D$ is the natural division ring of $R$, $K = \text{Cen} D$, and $A$ is a square-free $K$-algebra; that is, $\dim_K(eAf) \leq 1$ for every pair $e, f \in A$ of primitive idempotents.

1. Introduction.

Let $R$ be an artinian ring with basic ring $R$, and suppose that $R/J \cong D^n$ (as rings) for some division ring $D$ and some $n \in \mathbb{N}$. Then $R$ is said to be split with division ring $D$. If $K = \text{Cen} D$, then $R$ is a $D$-algebra in case also $R \cong D \otimes_K A$ for some (necessarily split) $K$-algebra $A$. Much of the study of such a $D$-algebra reduces to that of the usually more tractable $K$-algebra $A$. Of course, it is not common for an artinian ring $R$ to be split and even less common for it to be a $D$-algebra. Nevertheless, it is of some interest to identify classes of artinian rings that can be realized as $D$-algebras. Fortunately, as we show in Theorem 3.2, the property of being a $D$-algebra is Morita invariant, and so we may focus on basic rings.

An artinian ring $R$ is square-free in case no indecomposable projective $R$-module has a repeated composition factor. In [6] D’Ambrosia has shown that every indecomposable square-free ring is split. Moreover, there it is shown that there exist substantial classes of indecomposable square-free rings that are not $D$-algebras for any division ring $D$. In this paper we find a large class of square-free rings that are necessarily $D$-algebras.

Each artinian ring $R$ determines two digraphs, the quivers of $R$. (See, for example, [7].) If $R$ is square-free, then these two quivers are order-theoretic duals (see [6] Lemma 2.1), so we work with the right quiver $Q$ of $R$. We say that $Q$ is solvable if it contains no $n$-crown (see Section 2). So, for

1In [6], D’Ambrosia inadvertently reversed the roles of left and right quivers. The reader should replace “left quiver” with “right quiver” and vice versa when referring to that paper.
example, if $Q$ is a tree or has a unique source or a unique sink, then $Q$ is solvable.

Fuller and Haack ([7]) studied rings $R$ whose quivers are trees. Such rings are necessarily square-free and solvable. The main result of [7] is that, in effect, if the quivers of $R$ are trees, then $R$ is a $D$-algebra and its basic ring is isomorphic to a factor ring of the $D$-incidence ring of a poset.

In this paper we consider the more general square-free rings with solvable quivers. Let $Q$ be the quiver of a square-free ring $R$. If all cut edges of $Q$, viewed as an undirected graph, are removed, the connected components of the resulting graph are the cyclic components of $Q$. For the solvability of $Q$ to be reflected in the behavior of $R$ we need that $R$ be transitive on each of these cyclic components; we say then that $R$ is locally transitive. (See Section 4.) So if $Q$ is a tree, then it is locally transitive. Our main result, Theorem 4.8, is that if $R$ is an indecomposable, locally transitive, square-free ring with solvable quiver, then $R$ is a $D$-algebra.

Finally, in Section 5 we look at a generalization of Gabriel’s Theorem for split $K$-algebras. We show that a $D$-algebra is a factor of the $D$-path algebra of its quiver. From this it follows that, if $R$ is basic square-free locally transitive with solvable quiver, then $R$ is a twisted factor ring of the $D$-incidence algebra of its quiver.

2. Solvable quivers.

By a quiver we will mean a finite digraph $Q = (V, A)$ with vertex set $V$ and arrow set $A$. By a $d$-path (or directed path) in $Q$ we mean a sequence of arrows

$$p = (a_1, a_2, \ldots, a_m)$$

such that for each $1 \leq i < m$ the terminal vertex of $a_i$ is the initial vertex of $a_{i+1}$. A $d$-path with $m$ arrows has length $m$. A single vertex $v$ will be considered as a degenerate (trivial) $d$-path of length 0 with initial and terminal vertices both $v$. A bi-directed cycle in $Q$ is the union $p \cup q$ of two $d$-paths of $Q$ that intersect only in a common initial vertex and common terminal vertex.

We will also want to view the pair $(V, A)$ as an undirected graph; that is, in this context we consider each $a \in A$ as an unoriented edge. Then by a path in $Q$ we mean a path in the undirected graph $(V, A)$. A cycle in $Q$ is just a nontrivial path that is simply closed. Thus, every $d$-path of $Q$ is a path of $Q$, and every bi-directed cycle of $Q$ is a cycle of $Q$ or a union of cycles such as two loops with the same vertex. But the converses of these statements are true only in degenerate examples.
Example 2.1. Consider the following quivers:

\[ Q_1 \quad Q_2 \quad Q_3 \]

Here \( Q_1 \) has one cycle and one bi-directed cycle, the union of the two \( d \)-paths \((a, b)\) and \((c, d)\). The quiver \( Q_2 \) has one cycle and no bi-directed cycles. The graph \( Q_3 \) has two bi-directed cycles and one other cycle.

Let \( Q = (V, A) \) be a quiver that contains no directed cycles and hence no loops. A vertex is a **source** if it is the head of no arrow and it is a **sink** if it is the tail of no arrow. Since our quiver is finite and contains no directed cycles, it must have at least one source and at least one sink. Let \( U = \{u_1, \ldots, u_n\} \) be a set of sources and \( W = \{w_1, \ldots, w_n\} \) a set of sinks in \( Q \). If \( n \geq 2 \), then \( U \) and \( W \) determine an **\( n \)-crown** if:

\( (C1) \) There is a \( d \) path from \( u_i \) to \( w_j \) iff \( i > 1 \) and \( i - 1 \leq j \leq i \), or \( (i, j) = (1, 1) \), or \( (i, j) = (1, n) \).

\( (C2) \) If \( n = 2 \), no path from \( u_1 \) to \( w_2 \) shares a vertex with any path starting at \( u_2 \) and ending at \( w_1 \).

So, for example, the quiver \( Q_2 \) of Example 2.1 is a 2-crown. For \( n > 2 \) an \( n \)-crown has the form

\[ w_1 \quad w_2 \quad \cdots \quad w_{n-1} \quad w_n \]

\[ u_1 \quad u_2 \quad \cdots \quad u_{n-1} \quad u_n \]

where the arrows indicate directed paths of unspecified length. We say the quiver \( Q \) is **solvable** in case it contains no \( n \)-crown. So note that if \( Q \) has either a single source or a single sink, then \( Q \) is solvable.

Let \( U = \{u_1, \ldots, u_n\} \) be a set of sources and \( W = \{w_1, \ldots, w_n\} \) a set of sinks in \( Q \). If there are \( d \)-paths from \( u_i \) to \( w_i \) and \( w_{i-1} \) for all \( 1 < i \leq n \) and from \( u_1 \) to \( w_1 \) and to \( w_n \), then we say that \( U \) and \( W \) determine a **weak \( n \)-crown**. So an \( n \)-crown is a weak \( n \)-crown but not conversely.
Lemma 2.2. Suppose that $Q$ is a connected solvable quiver. Let $U = \{u_1, \ldots, u_n\}$ be a set of sources and $W = \{w_1, \ldots, w_n\}$ a set of sinks in $Q$. If $U$ and $W$ determine a weak $n$-crown, then there exist $1 < i \leq n$, $1 \leq j < n$ and a vertex $v$ with $d$-paths from $u_1$ to $w_n$ through $v$ and from $u_i$ to $w_j$ through $v$.

Proof. We’ll prove this by induction. For $n = 2$ this immediate from (C2). So let $n > 2$. Since $Q$ has no $n$-crowns, there is a $d$-path from some $u_i$ to some $w_j$ where

$$0 < j - i < n - 1 \text{ or } 1 < i - j < n.$$

In the first case

$$U' = \{u_1, \ldots, u_i, u_{j+1}, \ldots, u_n\} \text{ and } W' = \{w_1, \ldots, w_{i-1}, w_j, \ldots, w_n\}$$

form a weak $n - j + i$-crown. In the second,

$$U' = \{u_1, \ldots, u_j, u_i, \ldots, u_n\} \text{ and } W' = \{w_1, \ldots, w_j, w_i, \ldots, w_n\}$$

form a weak $n - i + j + 1$-crown. So by induction there exist the desired vertex $v$ and $d$-paths. □

Now let $Q = (V, A)$ be a quiver. By a subquiver of a quiver $Q$ we mean a quiver whose vertex and arrow sets are subsets of $V$ and $A$. If $H \subseteq A$, then we may view $H$ as a subquiver with arrow set $H$. If $H$ and $K$ are two subquivers of $Q$, then $H \cup K$ is the subquiver whose vertex and arrow sets are the unions of those of $H$ and $K$, and $H \cap K$ is the subquiver whose vertex and arrow sets are the intersections of those of $H$ and $K$. Also, if $H \subseteq A$, then the subgraph generated by $H$ is the (unoriented) subgraph of $Q$ consisting of all edges from $H$ together with the vertices of these edges. The complement of $H$ is the subgraph of $Q$ generated by the edges of $Q$ not in $H$.

Let $T$ be the subgraph generated by the set of all cut edges of $Q$. (A cut edge is an edge whose deletion separates the graph.) Let $S$ be the complement of $T$, so each edge of $S$ belongs to at least one cycle. Let $S_1, \ldots, S_r$ be the connected components of $S$; we call these the cyclic components of $Q$. The subgraph $T$ consists of a finite set of disjoint trees $T_1, \ldots, T_q$. Since each edge of $T$ is a cut edge of $Q$, each tree $T_i$ meets any one cyclic component $S_j$ in at most one common vertex. And if $Q$ is connected with $r \geq 1$, each tree $T_i$ meets at least one $S_j$ in a common vertex. Note that if $Q$ is connected and we shrink each cyclic component $S_i$ to a single vertex, then the resulting graph will be a tree.

Suppose that $Q$ is a finite quiver with no directed cycles. The addition or removal of a single arrow that is a cut edge of the graph will not affect the existence of an $n$-crown. Thus, we see that $Q$ is solvable iff each of its cyclic components is solvable.
For every cyclic component $S_i$, its **star** is the subgraph $\text{St}(S_i)$ of $Q$ generated by $S_i$ and all trees $T_j$ in $T$ that meet $S_i$. A labeling $\{S_1, \ldots, S_r\}$ of the components of $S$ is **proper order** in case for each $1 \leq k \leq r$ the graph

$$\text{St}(S_1) \cup \text{St}(S_2) \cup \cdots \cup \text{St}(S_k)$$

is connected.

**Lemma 2.3.** If $Q$ is a connected quiver, then $Q$ is a tree or there exists a proper order for the cyclic components of $Q$.

**Proof.** Assume that $Q$ is not a tree. Let $r$ be the number of cyclic components $S_1, \ldots, S_r$, and let $1 \leq k \leq r$. We induct on $k$. If $k = 1$, then the result is trivial. So let $1 \leq k < r$, and assume that cyclic components $S_1, \ldots, S_k$ have been chosen so that $H = \text{St}(S_1) \cup \text{St}(S_2) \cup \cdots \cup \text{St}(S_k)$ is connected. Since $Q$ is connected, the complement of $H$ must have at least one vertex in common with $H$. If some tree $T_j$ has a vertex in common with $H$, then it lies in $H$. Thus, there must be some cyclic component $S_{k+1}$ sharing a vertex with $H$. So, clearly, $H \cup \text{St}(S_{k+1})$ is connected. \qed

### 3. Square-free $D$-algebras.

Let $R$ be an artinian ring with radical $J$ and with basic set $E = \{e_1, \ldots, e_n\}$; that is, the set $\{e_1, \ldots, e_n\}$ consists of pairwise orthogonal primitive idempotents, and $Re_1, \ldots, Re_n$ is a complete irredundant set of the indecomposable projective left $R$-modules. (See [3], page 305.) Then (see [3], Exercise 32.14, or [7]) the (right) **quiver** of $R$ is the digraph $Q(R)$ with vertex set $E$ and with $m_{ij}$ arrows from $e_i$ to $e_j$ where $m_{ij}$ is the multiplicity of the simple module $e_jR = e_jR/e_jJ$ in the semisimple module $e_iJ/e_iJ^2$. The left quiver of $R$ is defined similarly.

An artinian ring $R$ is **square-free** in case for every primitive idempotent $e \in R$, the indecomposable projective modules $Re$ and $eR$ have no repeated composition factors. (See [6] for the basic facts about these rings.) If $R$ is square-free, then ([6], Lemma 2.1) the left quiver of $R$ is just the order-theoretic dual, $Q(R)^{op}$, of the right quiver $Q(R)$. So, in particular, for a square-free ring the left quiver is solvable iff the right quiver is solvable. For the remainder of this paper we will assume that $R$ is square-free. Note that this means that the quiver $Q$ of $R$ has no multiple arrows, and no arrow from a vertex to itself.

For a special case of such rings, let $A$ be a split finite dimensional algebra over a field $K$. Then the $K$-algebra $A$ is square free iff for each pair $e, f \in A$ of primitive idempotents

$$\dim_K(eAf) \leq 1.$$ 

These algebras have been described in detail in [2]. In particular, every incidence algebra of a pre-ordered set is square-free. Observe that if $A$ is a
square-free $K$-algebra and if $D$ is a division ring with $\text{Cen } D = K$, then

$$R = D \otimes_K A$$

is a square-free ring. A square-free ring of this form is said to be a square-free $D$-algebra. As is shown in [6], there are plenty of square-free rings that are not $D$-algebras. Let $R$ be an indecomposable square-free ring. Our main goal in this paper is to show that if $R$ is locally transitive square-free (see Section 4) with solvable quiver, then it must be a $D$-algebra for some $D$. Our intermediate goal in this section is to obtain a characterization of square-free $D$-algebras that will provide the machinery to prove the main result.

So now let $R$ be an indecomposable square-free ring. Then (see [6], Theorem 1.4) there is a division ring $D$, called the division ring of $R$, such that for every primitive idempotent $e \in R$, $D \cong eRe$. Moreover, for each pair $e, f \in R$ of primitive idempotents, $\dim_{eRe}(eRf) \leq 1$ and $\dim_{fRf}(eRf) \leq 1$. (See [6], Theorem 1.4.) It follows ([6], Corollary 1.5) that if $e, f \in R$ are primitive idempotents with $eRf \neq 0$, then for each $0 \neq erf \in eRf$, there is an isomorphism

$$\varphi_{erf} : eRe \rightarrow fRf,$$

that we treat as a right operator \(^2\) where $(exe)\varphi_{erf}$ is the unique element of $fRf$ such that

$$exe \cdot erf = erf \cdot (exe)\varphi_{erf}.$$  

This is the inverse of the isomorphism given in [6]. Concerning the isomorphisms $\varphi_{erf}$, we will use the following slight variation of Corollary 1.5 from [6]:

**Lemma 3.1.** Let $e, f, g \in R$ be primitive idempotents in the square-free ring $R$ and let $r, s, t \in R$. If $erfsg = (eae)etg \neq 0$ for some $eae \in \text{Cen}(eRe)$, then

$$\varphi_{erf}\varphi_{fsg} = \varphi_{etg}.$$  

**Proof.** For each $exe \in eRe$,

$$(etg) \cdot (exe)\varphi_{erf}\varphi_{fsg} = (eae)^{-1}erfsg \cdot (exe)\varphi_{erf}\varphi_{fsg}$$

$$= (eae)^{-1}erf \cdot (exe)\varphi_{erf} \cdot fsg$$

$$= (eae)^{-1} \cdot exe \cdot erfsg$$

$$= (exe)(eae)^{-1}erfsg$$

$$= (exe)(etg)$$

$$= (etg)(exe)\varphi_{etg}.$$  

And so $\varphi_{erf}\varphi_{fsg} = \varphi_{etg}$. \(\square\)

\(^2\)In this section and the next all maps will be written as right operators unless explicitly stated otherwise.
Next, we observe that the property of being a $D$-algebra is Morita invariant, so that we can reduce to the basic case.

**Theorem 3.2.** Let $R$ and $S$ be Morita equivalent rings. If $R \cong D \otimes_K A$ is a $D$-algebra, then there is a $K$-algebra $A'$, Morita equivalent to $A$, such that $S \cong D \otimes_K A'$.

**Proof.** Suppose that $R = D \otimes_K A$. If $e, f \in A$ are primitive idempotents with $Ae \cong Af$, then $1 \otimes e$ and $1 \otimes f$ are primitive idempotents of $R$ that satisfy $R(1 \otimes e) \cong R(1 \otimes f)$. Thus, $\{e_1, \ldots, e_n\}$ is a basic set of primitive idempotents for $A$ iff $\{1 \otimes e_1, \ldots, 1 \otimes e_n\}$ is a basic set for $R$. Assuming that those sets are basic, then

$$P' = \bigoplus_{i=1}^{n} (R(1 \otimes e_i))^{m_i} = D \otimes_K \bigoplus_{i=1}^{n} (Ae_i)^{m_i}$$

is a progenerator for $R$ iff $m_i \geq 1$ for each $i$, iff $P = \bigoplus_{i=1}^{n} (Ae_i)^{m_i}$ is a progenerator for $A$. So it will suffice to prove that if $AP$ is a finitely generated projective $A$-module, then (see [3], Corollary 22.4)

$$\text{End}_{R}(D \otimes_K P) \cong D \otimes_K \text{End}_{A}(P).$$

For in this case, we let $A' = \text{End}_{A}(P)$, where $P$ is a progenerator for $A$ as described above.

But if $P = A^n$ is free, then

$$\text{End}_{R}(D \otimes_K A^n) \cong \text{End}_{R}(R^n)$$

$$\cong M_n(R)$$

$$\cong D \otimes_K M_n(A)$$

$$\cong D \otimes_K \text{End}_{A}(A^n).$$

Finally, if $e' \in \text{End}_{A}(A^n)$ is an idempotent endomorphism, then there is an idempotent $e \in \text{End}_{R}(R^n)$ with

$$\text{End}_{R}(D \otimes_K A^n e') \cong \text{End}_{R}(R^n e)$$

$$\cong e \text{End}_{R}(R^n) e$$

$$\cong (1 \otimes e')(D \otimes_K \text{End}_{A}(A^n))(1 \otimes e')$$

$$\cong D \otimes_K e' \text{End}_{A}(A^n)e'$$

$$\cong D \otimes_K \text{End}_{A}(A^n e'),$$

as claimed. \qed

For the remainder of this section, we will assume that $R$ is a basic indecomposable square-free ring with complete set $E = \{e_1, \ldots, e_n\}$ of orthogonal primitive idempotents. We define a binary relation $\leq$ on $E$ by

$$e_i \leq e_j \quad \text{in case} \quad e_iRe_j \neq 0.$$
In general, this relation need not be either transitive or anti-symmetric.

For each \( e_i \leq e_j \) in \( E \), fix \( 0 \neq s_{ij} \in e_iRe_j \) with the proviso that \( s_{ii} = e_i \) for each \( i = 1, \ldots, n \). Then the set

\[ B = \{ s_{ij} : e_i \leq e_j \} \]

is a basis for \( R \) (over \( E \)). So if \( B = \{ s_{ij} \} \) is a basis for \( R \) over \( E \), then since \( \dim_{e_iRe_i}(e_iRe_j) \leq 1 \) for each \( i, j \), we have that as additive groups,

\[ R = \sum_{i=1}^{n} \sum_{j=1}^{n} e_iRe_j = \sum_{e_i \leq e_j} e_iRe_i s_{ij}. \]

If \( B \) is a basis for \( R \) over \( E \), then we will usually abbreviate the isomorphisms \( \varphi_{s_{ij}} \) by

\[ \varphi_{ij} = \varphi_{s_{ij}} : e_iRe_i \rightarrow e_jRe_j \]

for each \( s_{ij} \in B \).

Let \( B \) be a basis for \( R \) over \( E \) and let \( B' \subseteq B \) be a subset of \( B \). We say that \( B' \) is solvable in case for each \( i = 1, \ldots, n \) there is an isomorphism \( \mu_i : D \rightarrow e_iRe_i \) such that

\[ \varphi_{ij} = \mu_i^{-1} \mu_j \]

for all \( s_{ij} \in B' \). We will also then say that \( (\mu_i)_{i=1}^{n} \) is a solution on \( B' \).

In particular, the basis \( B \) is a solvable basis for \( R \) if there is a solution \( (\mu_i)_{i=1}^{n} \) on \( B \). Note that if \( (\mu_i)_{i=1}^{n} \) is a solution on \( B' \) and if \( \alpha \in \text{Aut} D \), then \( (\alpha \mu_i)_{i=1}^{n} \) is also a solution on \( B' \).

**Lemma 3.3.** Let \( B' \subseteq B \) be solvable. For all \( e_i \leq e_j \leq e_k \) in \( E \) with \( e_i \leq e_k \), if \( s_{ij}, s_{jk}, s_{ik} \in B' \), then there is a (necessarily unique) \( d = d_{ijk} \in \text{Cen}(e_iRe_i) \) with

\[ s_{ij}s_{jk} = ds_{ik}. \]

**Proof.** Since \( s_{ij}s_{jk} \in e_iRe_i s_{ik} \), there is a (necessarily unique) \( d \in e_iRe_i \) with \( s_{ij}s_{jk} = ds_{ik} \). We claim that \( d \in \text{Cen}(e_iRe_i) \). By hypothesis, there exist isomorphisms \( \mu_h : D \rightarrow e_hRe_h \) such that \( \varphi_{gh} = \mu_g^{-1} \mu_h \) for all \( s_{gh} \in B' \). Thus, for each \( d' \in e_iRe_i \),

\[
\begin{align*}
  d'ds_{ik} &= d's_{ij}s_{jk} \\
  &= s_{ij}s_{jk}(d')\varphi_{ij}\varphi_{jk} \\
  &= s_{ij}s_{jk}(d')\mu_i^{-1}\mu_k \\
  &= ds_{ik}(d')\varphi_{ik} \\
  &= dd's_{ik}.
\end{align*}
\]

Thus, \( d'd = dd' \) and \( d \in \text{Cen}(e_iRe_i) \), as claimed. \( \square \)

**Corollary 3.4.** Let \( B' \subseteq B \) be solvable, and let \( e_i \leq e_j \leq e_k \) in \( E \) with \( e_i \leq e_k \). If \( s_{ij}, s_{jk}, s_{ik} \in B' \), then \( \varphi_{ik} = \varphi_{ij}\varphi_{jk} \).

**Proof.** Apply Lemmas 3.1 and 3.3. \( \square \)
We now characterize square-free $D$-algebras in terms of solvable bases. For a different characterization, see Lemma 3.1 of [6].

**Theorem 3.5.** Let $R$ be a basic indecomposable square-free ring with division ring $D$. Then $R$ is a $D$-algebra if and only if $R$ has a solvable basis.

**Proof.** ($\Rightarrow$) Let $R = D \otimes_K A$ for some square-free $K$-algebra $A$ where $K = \text{Cen} D$. We may assume that $E = \{1 \otimes e_i : e_i \in E_A\}$ is a complete set of orthogonal primitive idempotents for $R$ where $E_A = \{e_1, \ldots, e_n\}$ is a complete set of orthogonal primitive idempotents for $A$. We may choose a basis $B_A = \{s_{ij} : e_i \leq e_j\}$ for $A$ over $E_A$; then $B = \{1 \otimes s_{ij} : e_i \leq e_j\}$ is a basis for $R$ over $E$. For each $i$, let $\mu_i : d \mapsto d \otimes e_i$. Then since $e_i Ae_i \cong K$ for each $i$, each $\mu_i : D \mapsto (1 \otimes e_i)R(1 \otimes e_i)$ is an isomorphism. For each $s_{ij} \in B_A$ and $d \in D$,

$$(d \otimes e_i)(1 \otimes s_{ij}) = (d \otimes e_is_{ij}) = (d \otimes s_{ij}e_i) = (1 \otimes s_{ij})(d \otimes e_j).$$

So $(d \otimes e_i)\varphi_{ij} = d \otimes e_j$, and $(\mu_i)_{i=1}^n$ is a solution for $B$. Thus, $B$ is a solvable basis for $R$.

($\Leftarrow$) Let $E = \{e_1, \ldots, e_n\}$ be a complete set of orthogonal primitive idempotents for $R$ and let $B = \{s_{ij} : e_i \leq e_j\}$ be a basis for $R$. Say $(\mu_i)_{i=1}^n$ is a solution for $B$. Define $\psi : D \mapsto R$ by

$$(d)\psi = \sum_{i=1}^n (d)\mu_i.$$ 

Observe that for all $e_i \leq e_j$ in $E$ and all $r \in R$,

$$(d)\psi \cdot e_i re_j = (d)\mu_i \cdot e_i re_j \quad \text{and} \quad e_i re_j \cdot (d)\psi = e_i re_j \cdot (d)\mu_j.$$ 

The map $\psi : D \mapsto R$ is clearly additive and since $(d)\mu_i (d')\mu_j = \delta_{ij} (dd')\mu_i$, where $\delta_{ij}$ is the Kronecker delta, it follows that $\psi$ is an isomorphism from $D$ onto a division subring $D^\psi$ of $R$.

We claim next that if $K = \text{Cen} D$, then $K^\psi = (K)\psi$ is the center of $R$. First, we note that for each $i$, we have $(K)\mu_i = \text{Cen}(e_i Re_i)$. Now suppose that $r \in \text{Cen}(R)$. Then for each $i$, we have $e_i r = re_i$, so

$$r = e_1 re_1 + \cdots + e_n re_n.$$ 

Clearly, $e_i re_i \in \text{Cen}(e_i Re_i) = (K)\mu_i$. Next, observe that if $e_i Re_j \neq 0$, then

$$(e_i re_i)s_{ij} = s_{ij}(e_j re_j),$$ 

so $e_j re_j = (e_i re_i)\varphi_{ij}$ or

$$(e_j re_j)\mu_j^{-1} = (e_i re_i)\mu_i^{-1} \in K.$$ 

Thus, there exists some unique $k \in K$ with $(k)\mu_i = e_i re_i$ for all $i = 1, \ldots, n$. That is, $r = (k)\psi \in K^\psi$. 
On the other hand, let \( d \in K \). Then certainly, each \( (d)\mu_i \in \text{Cen}(e_iRe_i) \).
So it will suffice to show that for each \( i, j \) with \( e_iRe_j \neq 0 \), and for each \( r \in R \),
\[
(d)\mu_i \cdot (e_i re_j) = (e_i re_j) \cdot (d)\mu_j.
\]
But if \( e_i re_j \neq 0 \), then \( e_i re_j = e_i r'e_i s_{ij} \) for some \( e_i r'e_i \in e_i Re_i \). Thus, since
\( (d)\mu_i \in \text{Cen}(e_i Re_i) \),
\[
(d)\mu_i \cdot (e_i re_j) = (e_i r'e_i) \cdot (d)\mu_i \cdot s_{ij}
\]
\[
= (e_i r'e_i) s_{ij} \cdot (d)\mu_i \mu_j
\]
\[
= (e_i r'e_i) s_{ij} \cdot (d)\mu_i \mu_j^{-1} \mu_j
\]
\[
= (e_i re_j) \cdot (d)\mu_j.
\]

Thus, \( \text{Cen}(R) = K^\psi \) as claimed.

Now let \( A \) be the \( K^\psi \) subspace of \( R \) spanned by \( B \). By Lemma 3.3, for each \( s_{ij}, s_{jk} \in B \), if \( s_{ij} s_{jk} \neq 0 \), then
\[
s_{ij} s_{jk} = ds_{ik}
\]
for some \( d \in \text{Cen}(e_i Re_i) \). But \( \text{Cen}(e_i Re_i) = (K)\mu_i \), so there is a (necessarily unique) \( t \in K \) with \( (t)\mu_i = d \). Then \( (t)^\psi \in \text{Cen}(R) \) and
\[
s_{ij} s_{jk} = (t)^\psi s_{ik}.
\]
Thus, \( A \) is actually a \( K^\psi \)-subalgebra of \( R \). In fact, it is square-free. Indeed, \( E \) is a complete set of idempotents for \( A \), and for each \( e_i, e_j \), if \( e_i Ae_j \neq 0 \), then \( e_i Ae_j \) is one dimensional over \( K^\psi \) generated by \( s_{ij} \).

Next, we claim that for each \( a \in A \) and each \( d \in D \),
\[
a(d)^\psi = (d)^\psi a.
\]
For this, it will clearly suffice to assume that \( a = s_{ij} \) for some \( e_i Re_j \neq 0 \). But
\[
((d)^\psi)(s_{ij}) = ((d)\mu_i)(s_{ij})
\]
\[
= (s_{ij})((d)\mu_i \varphi_{ij})
\]
\[
= (s_{ij})((d)^\psi)
\]
so \( a(d)^\psi = (d)^\psi a \), as claimed.

Finally, the map \( \tau : D \times A \rightarrow R \) defined by
\[
\tau : (d, a) \mapsto ((d)^\psi)a
\]
is clearly \( K \)-bilinear and universal, so there is the desired ring isomorphism \( \tau : D \otimes_K A \rightarrow R \). \qed
4. The main result.

Let $R$ be a basic square-free ring with basic set $E = \{e_1, \ldots, e_t\}$ and with quiver $Q = Q(R)$. In this section we shall show that under certain conditions a solvable quiver will imply that $R$ is a $D$-algebra. Since we want the path structure of $Q$ to be reflected in the algebra of $R$, we impose a restriction on $R$ that will provide the desired connection between $Q$ and $R$.

If $E$ is some set of vertices in $Q$, then we say that $R$ is transitive on $E$ in case for all $e, f, g \in E$

$$eJfJg = 0 \implies eJf = 0 \text{ or } fJg = 0.$$  

We say that $R$ is transitive if it is transitive on the set $E$ of all vertices of $Q$ and locally transitive if it is transitive on the vertices of each cyclic component of $Q$. If $R$ is transitive, then (see \[6\]) the relation $\leq$ on $E$, defined in Section 3, is a partial order, and $Q$ is the Hasse diagram for this poset. For more on the property of transitivity for square-free rings see \[6\]. Much of the significance of transitivity for us is given in the next pair of lemmas.

**Lemma 4.1.** Let $R$ be transitive. For each pair $e_i, e_j \in E$, the following are equivalent:

(a) $e_i \leq e_j$;
(b) $e_iRe_j \neq 0$;
(c) There is a $d$-path from $e_i$ to $e_j$ in $Q$.

In particular, $e_{i_0} \leq e_{i_1} \leq \cdots \leq e_{i_k}$ in $Q$ if there is a $d$-path in $Q$ through the sequence $(e_{i_0}, e_{i_1}, \ldots, e_{i_k})$.

**Proof.** The equivalence of (a) and (b) is by definition.

(b) $\iff$ (c). If there is an arrow from $e_i$ to $e_k$ and an arrow from $e_k$ to $e_j$, then $e_iJRe_k \neq 0$ and $e_kJe_j \neq 0$, so by transitivity, $e_iJe_kJe_j \neq 0$. The conclusion follows by induction on the length of the path.

(b) $\implies$ (c). We may assume $e_i \neq e_j$. If $e_iRe_j \neq 0$, then $e_iJ^m e_j \neq 0$ for some maximal $m \geq 1$. If $m = 1$, then there is an arrow from $e_i$ to $e_j$. If $m > 1$, then there is an $e_k$ with $e_iJ^m e_j \supseteq e_iJe_kJ^{m-1}e_j \neq 0$. So there is an arrow from $e_i$ to $e_k$ and, by induction, a $d$-path from $e_k$ to $e_j$.  

Observe that if $R$ is transitive, then $Q$ contains no directed cycles. Let

$$B = \{s_{ij} : e_i \leq e_j\}$$

be a basis for $R$ over $E$.

**Lemma 4.2.** If $R$ is transitive, and if there is a $d$-path in $Q$ through the sequence $(e_{i_0}, e_{i_1}, \ldots, e_{i_k})$ of vertices in $E$, then $s_{i_0i_1} \cdots s_{i_{k-1}i_k} \neq 0$. 

Proof. Suppose that \( e_1, e_2, e_3 \in E \) and that there is a \( d \)-path through the sequence \((e_1, e_2, e_3)\). We may assume that \( e_1, e_2, e_3 \) are distinct. Then by Lemma 4.1, \( e_1Re_2 \neq 0 \) and \( e_2Re_3 \neq 0 \). So
\[
e_1Je_2 = e_1Re_2 = e_1Re_1s_{12} \neq 0,
\]
and similarly, \( e_2Je_3 = s_{23}e_3Re_3 \neq 0 \). Thus, by transitivity, \( e_1Je_2Je_3 \neq 0 \), so that \( s_{12}s_{23} \neq 0 \). Now induct.

Lemma 4.3. If \( R \) is transitive, and if \( e_i \leq e_j \leq e_k \leq e_\ell \) in \( E \), then there exists \( v \in e_\ell Re_k \) with
\[
s_{i\ell} = s_{ij}us_{k\ell}.
\]
Proof. This follows immediately from the fact that since \( R \) is transitive and square-free, \( e_iRe_\ell = e_iRe_\ell Re_kRe_\ell = s_{ij}e_\ell Re_kRe_\ell \). See [6].

Our next goal is to find sufficient conditions on \( R \) for it to have a solvable basis. For this we first assume that \( R \) is transitive and indecomposable. We will prove the main result for this case, extend that to stars of cyclic components, and then piece these together for the general result. A key part of that is a lemma that gives a condition under which we can extend solvability from a subquiver to a larger subquiver.

Let \( \mathcal{Q}_1 \) be a subquiver of \( \mathcal{Q} \). If \( B \) is a basis for \( R \), then we set
\[
B(\mathcal{Q}_1) = \{s_{ij} \in B : \text{there is a \( d \)-path from } e_i \text{ to } e_j \text{ in } \mathcal{Q}_1\}.
\]
Let \( V \subseteq E \). We denote by \( V^+ \) the subquiver of \( \mathcal{Q} \) whose arrows are precisely those that lie on some \( d \)-path of \( \mathcal{Q} \) that originates at some \( e \in V \). Similarly, we denote by \( V^- \) the subquiver whose arrows lie on some \( d \)-path of \( \mathcal{Q} \) that terminates at some \( e \in V \). So, the vertex set of \( V^+ \) is the subposet of \( \mathcal{Q} \) consisting of all upper bounds of the elements of \( V \) and the vertex set of \( V^- \) is the subposet of all lower bounds of the elements of \( V \).

Next, we want to investigate some conditions under which a basis solvable on a subquiver can be modified to create a basis solvable on a larger subquiver.

Lemma 4.4. Let \( R \) be an indecomposable square-free ring with quiver \( \mathcal{Q} \), and suppose that \( B \) is a basis for \( R \). Let \( \mathcal{Q}_1 \) and \( \mathcal{Q}_2 \) be disjoint subquivers with both \( B(\mathcal{Q}_1) \) and \( B(\mathcal{Q}_2) \) solvable. If there is an arrow \( a \) from a vertex \( e_i \) in \( \mathcal{Q}_1 \) to a vertex \( e_j \) in \( \mathcal{Q}_2 \), then there is a basis \( B' \) with \( B'(\mathcal{Q}_1 \cup \{a\} \cup \mathcal{Q}_2) \) solvable. In particular, if a subquiver \( \mathcal{Q}' \) of \( \mathcal{Q} \) is a tree, then there is a basis \( B' \) for \( R \) with \( B'(\mathcal{Q}') \) solvable.

Proof. Let \((\mu_k)_{k=1}^n\) be a sequence of isomorphisms \( \mu_k : \rightarrow e_kRe_k \) that form solutions on both \( B(\mathcal{Q}_1) \) and \( B(\mathcal{Q}_2) \). For each \( e_h \leq e_k \) in \( \mathcal{Q} \) let \( s'_{hk} \in e_hRe_k \) be given by
\[
s'_{hk} = \begin{cases} 
  s_{hj}s_{jk}, & \text{if } e_h \in \mathcal{Q}_1 \text{ and } e_k \in \mathcal{Q}_2; \\
  s_{hk}, & \text{otherwise}.
\end{cases}
\]
Then \( B' = \{ s'_{hk} : e_h \leq e_k \} \) is a basis for \( R \). Moreover, the sequence \((\nu_k)_{k=1}^n \) of isomorphisms given by

\[
\nu_k = \begin{cases} 
\mu_j \varphi_{s_{kj}}^{-1}, & \text{if } e_k \in \mathcal{Q}_1 \text{ and } e_k \leq e_j; \\
\mu_k, & \text{otherwise}
\end{cases}
\]

is a solution for \( B'(\mathcal{Q}_1 \cup \{a\} \cup \mathcal{Q}_2) \). The final statement is now a simple induction. \( \square \)

**Lemma 4.5.** Let \( R \) be a transitive square-free ring with quiver \( \mathcal{Q} \) and basis \( B \). Let \( \mathcal{Q}_1 \) be a connected subquiver of \( \mathcal{Q} \), let \( a \) be an arrow of \( \mathcal{Q}_1 \), and let \( \mathcal{Q}_2 \) be the subquiver of \( \mathcal{Q}_1 \) obtained by deleting the arrow \( a \). Assume that \( B(\mathcal{Q}_2) \) is solvable. If \( a \) belongs to a bi-directed cycle of \( \mathcal{Q}_1 \), then there is a basis \( B' \) for \( R \) with \( B'(\mathcal{Q}_1) \) solvable.

**Proof.** Let \((\mu_i)_{i=1}^n \) be a sequence of isomorphisms \( \mu_i : D \rightarrow e_i Re_i \) that form a solution for \( B(\mathcal{Q}_2) \). Suppose that \( a \) is an arrow from \( e_i \) to \( e_j \) and belongs to the bi-directed cycle \( p \cup q \) where \( p \) and \( q \) are \( d \)-paths from \( e_i \) to \( e_m \). We may assume that \( a \) is an arrow on \( q \). By Lemma 4.3 there is a \( \varphi \in e_i Re_j \) with \( s_{km} = s_{ki} vs_{jm} \). Thus,

\[
\varphi \in \varphi_{s_{ki}}^{-1} \varphi_{s_{jm}}^{-1} = \mu_i^{-1} \mu_j.
\]

For each \( e_h \leq e_k \) in \( E \) define \( s'_{hk} \in e_h Re_k \) by

\[
s'_{hk} = \begin{cases} 
s_{hi} vs_{jk}, & \text{if } e_h \leq e_i \leq e_j \leq e_k; \\
s_{hk}, & \text{otherwise}.
\end{cases}
\]

Then \( B' = \{ s'_{hk} : e_h \leq e_k \text{ in } E \} \) is a basis for \( R \) and \((\mu_i)_{i=1}^n \) is a solution for \( B'(\mathcal{Q}_1) \). \( \square \)

Let \( X \) be an arbitrary quiver and let \( V \) be the \( \mathbb{Z}_2 \)-space on the set \( A \) of all arrows of \( X \). Thus, we can also think of \( V \) as the Boolean algebra \( \mathcal{P}(A) \) of all subsets of \( A \). In the proof of the next lemma we will consider the \( \mathbb{Z}_2 \)-subspace \( B(X) \) of \( V \) spanned by the bi-directed cycles of \( X \).

**Lemma 4.6.** Let \( H \) be a connected subquiver of \( \mathcal{Q} \), let \( u \) be a minimal element of \( \mathcal{Q} \), let \( w \) be a maximal element of \( \mathcal{Q} \), let \( p_0 \) be a \( d \)-path from \( u \) to \( w \), and let \( X \) be a subquiver of \( u^+ \cap w^- \) such that \( p_0 \) lies in \( X \) and every arrow in \( X \) lies on some \( d \)-path in \( X \) from \( u \) to \( w \). Suppose there is a basis \( B \) for \( R \) with \( B(H \cup p_0) \) solvable. If \( w \in H \) or if \( u \in H \), then there is a basis \( B' \) for \( R \) with \( B'(H \cup X) \) solvable.

**Proof.** We’ll prove the case with \( w \in H \); the other is dual. The proof will be by induction on \( b = \dim B(H \cup X) - \dim B(H \cup p_0) \), the difference between the dimensions of the bi-directed cycle spaces of \( H \cup X \) and \( H \cup p_0 \). So suppose that \( b = 0 \). If there were some arrow \( a \) in \( X \) but not in \( H \cup p_0 \), then since there must be a \( d \)-path in \( X \) from \( u \) to \( w \) through \( a \), and since \( a \)
cannot be on \( p_0 \), there is some bi-directed cycle in \( X \) containing \( a \), contrary to \( \dim \mathcal{B}(H \cup X) = \dim \mathcal{B}(H \cup p_0) \). So if \( b = 0 \), then \( H \cup X = H \cup p_0 \), and we’re done.

So assume that \( b > 0 \). Then there must be some nontrivial \( d \)-path in \( X \) no arrow of which is in \( H \cup p_0 \). Let \( p_1 \) be such a \( d \)-path of maximal length. Say the initial vertex of \( p_1 \) is \( x \) and the terminal vertex is \( y \). If there is a vertex of \( p_1 \) other than \( y \) that is the initial vertex of a bi-directed cycle in \( X \), let \( x' \) be the greatest such vertex. Otherwise let \( x' = x \). If there is a vertex different from \( x' \) on \( p_1 \) between \( x' \) and \( y \) that is the terminal vertex of a bi-directed path in \( X \), then let \( y' \) be the smallest such vertex. Otherwise, let \( y' = y \). Let \( p \) be the sub \( d \)-path of \( p_1 \) from \( x' \) to \( y' \), and let \( X' = X \setminus p \). Then \( p_0 \) is in \( X' \) and every arrow in \( X' \setminus H \) lies on a \( d \)-path from \( u \) to \( w \) in \( X' \). But clearly, \( \dim \mathcal{B}(H \cup X') < \dim \mathcal{B}(H \cup X) \), so by induction, there is a basis \( B' \) for \( R \) with \( B'(H \cup X') \) solvable. Now every vertex of \( p \) other than \( x' \) and \( y' \) has valence 2 in \( H \cup X \). Thus, by Lemma 4.4, there is a basis \( B'' \) for \( R \) with \( B''(H \cup X) \) solvable. \( \square \)

**Lemma 4.7.** Let \( R \) be an indecomposable transitive square-free ring with quiver \( Q \). If \( Q \) is solvable, then \( R \) has a solvable basis.

**Proof.** Since \( Q \) is connected, there exist sequences \((u_1, \ldots, u_m)\) and \((w_1, \ldots, w_n)\) of minimal and maximal elements of \( Q \), with repetitions allowed, such that

\[
u_i, u_{i+1} \leq w_i \quad \text{for all } i < n \quad \text{and} \quad u_n \leq w_n,
\]

and \( Q = \{u_1, \ldots, u_n\}^+ \cap \{w_1, \ldots, w_n\}^- \). We begin with some notation. For each \( m \) let

\[
F_m = \{u_1, \ldots, u_m\}^+ \cap \{w_1, \ldots, w_m\}^-,
\]

and

\[
G_m = \{u_1, \ldots, u_m\}^+ \cap \{w_1, \ldots, w_{m-1}\}^-.
\]

To prove this lemma, it suffices to show that for each \( m \):

(i) For \( 1 \leq m \leq n \), if \( B \) is a basis for \( R \) with \( B(G_m) \) solvable, then there is a basis \( B' \) for \( R \) with \( B'(F_m) \) solvable.

(ii) For \( 1 \leq m < n \), if \( B \) is a basis for \( R \) with \( B(F_m) \) solvable, then there is a basis \( B' \) for \( R \) with \( B'(G_{m+1}) \) solvable.

We will show (i); the arguments for (ii) are similar.

So let \( m = 1 \). Then there must be a \( d \)-path \( p_0 \) from \( u_1 \) to \( w_1 \). By Lemma 4.4 there is a basis \( B \) for \( R \) with \( B(p_0) \) solvable. Then by Lemma 4.6, there is a basis \( B' \) for \( R \) with \( B'(u_1^+ \cap w_1^-) \) solvable. Next, let \( 1 < m \leq n \) and assume that \( B(G_m) \) is solvable. For \( 0 \leq k < m \), set

\[
H_k = (\{u_{m-k}, \ldots, u_m\}^+ \cap w_m^-) \cup G_m.
\]

We will show by induction on \( k \) that there is a basis \( B' \) for \( R \) with \( B'(H_k) \) solvable. Let \( k = 0 \). Then there is a \( d \)-path \( p_0 \) from \( u_m \) to \( w_m \), so by
Lemmas 4.4 and 4.5 there is a basis $B''$ for $R$ with $B''(G_m \cup p_0)$ solvable. Thus, by Lemma 4.6 there is a basis $B'$ for $R$ with $B'(H_0)$ solvable. Now let $0 \leq k < m - 1$ and assume there is a basis $B$ for $R$ with $B(H_k)$ solvable. If $u_{m-k-1} \in \{u_{m-k}, \ldots, u_m\}$ or if there is no $d$-path from $u_{m-k-1}$ to $w_m$, then $H_{k+1} = H_k$, and we’re done. Otherwise, by hypothesis about the sequences $(u_1, \ldots, u_n)$ and $(w_1, \ldots, w_n)$, there must be sets $U' = \{u_{m-k-1} = u'_1, u'_2, \ldots, u'_k\} \subseteq \{u_{m-k-1}, \ldots, u_m\}$ and $W' = \{w'_1, w'_2, \ldots, w'_h = w_m\} \subseteq \{w_{m-k-1}, \ldots, w_m\}$ for which $U'$ and $W'$ form a weak $n$-crown. So since $Q$ is solvable, Lemma 2.2 guarantees a pair of $d$-paths $p_0$ from $u'_1$ to $w_h'$. Let $q$ from some $u'_i \in U' \setminus \{u'_1\}$ to some $w'_j \in W' \setminus \{w'_h\}$ that meet in some common vertex $v$. But $d$-paths from $u'_1$ to $w'_j$ and from $u'_i$ to $w'_h$ must be in $H_k$, so that $p_0$ is in $H_k$. In particular, $H_k \cup p_0 = H_k$, and so by Lemma 4.6, there is a basis $B'$ for $R$ such that $B'(H_{k+1})$ is solvable. By induction on $k$, then, there is a basis $B'$ for $R$ with $B'(H_{m-1})$ solvable. But $H_{m-1} = F_m$, so we are done with (i). \hfill \Box

**Theorem 4.8.** If $R$ is a locally transitive indecomposable square-free ring with solvable quiver, then $R$ has a solvable basis and hence is a $D$-algebra.

**Proof.** By Theorem 3.2 we may assume that $R$ is basic. By Lemma 2.3 we may assume that the cyclic components of the quiver $Q$ of $R$ are labeled in a proper order $\{S_1, \ldots, S_n\}$. By Lemma 4.7 for each $i = 1, \ldots, n$ there is a basis $B_i$ for $R$ with $B_i(S_i)$ solvable. For each tree $T_j$ in the subquiver of cut edges of $Q$ there is a basis solvable on $T_j$. So by Lemma 4.4, for each $i = 1, \ldots, n$, we may assume that $B_i(\text{St}(S_i))$ is solvable. Finally, use Lemma 4.4 and induction on $n$ to find a basis $B$ that is solvable on $\text{St}(S_1) \cup \text{St}(S_2) \cup \cdots \cup \text{St}(S_n) = Q$. By Theorem 3.5, $R$ is a $D$-algebra. \hfill \Box

5. Matrix representations.

If $R = D \otimes_K A$ is a $D$-algebra, then $R$ acts essentially like the split $K$-algebra $A$. For example, a $D$-automorphism of $R$ is just a ring automorphism $\alpha$ of $R$ such that for some $K$-automorphism $\overline{\alpha}$ of $A$ (with $\alpha$ and $\overline{\alpha}$ now viewed as left operators)

$$\alpha(d \otimes a) = d \otimes \overline{\alpha}(a),$$

for all $d \in D$ and $a \in A$. Thus, the $D$-automorphism groups of $R$ are completely determined by the $K$-automorphism groups of $A$. For square-free algebras these groups have been studied extensively. (See, for example, [2].) So if $R$ is a square-free ring satisfying the hypotheses of Theorem 4.8, then the entire analysis of the $K$-automorphism groups of the square-free algebra $A$ given in [2] translates to that of the $D$-automorphism groups of $R$.

In this section, however, we want to describe how the $D$-algebras $R$ that are locally transitive square-free rings with solvable quivers can be realized as twisted matrix rings. Thanks to Theorem 3.2 we can restrict ourselves to
basic rings. The extension of these representations to the more general case is straightforward.

First, though, let \( R = D \otimes_K A \) be an arbitrary basic square-free \( D \)-algebra. Then the \( K \)-algebra \( A \) is straightforward. The extension of these representations to the more general case by \( 1 \otimes e_1, \ldots, 1 \otimes e_n \) is a basic set for \( R \), and if \( J \) is the Jacobson radical for \( A \), then \( D \otimes J \) is the Jacobson radical for \( R \). In particular, both \( A \) and \( R \) have the same quiver \( \mathcal{Q} \), and \( R \) is transitive (locally transitive) iff \( A \) is.

Now let \( R = D \otimes_K A \) be a basic, locally transitive, indecomposable square-free ring with quiver \( \mathcal{Q} \) and division ring \( D \), and assume that \( E = \{ e_1, \ldots, e_n \} \) is a basic set of primitive idempotents for \( A \). Define a relation \( \preceq \) on \( E \) by

\[
e_i \preceq e_j \iff \text{there is a } d\text{-path in } \mathcal{Q} \text{ from } e_i \text{ to } e_j.
\]

Let \( e_i \) and \( e_j \) belong to the same cyclic component of \( \mathcal{Q} \). By Lemma 4.1, \( e_i \preceq e_j \) iff \( e_i A e_j \neq 0 \). Thus, by Theorem 2.3 of [6], restricted to each cyclic component, this relation is a partial order with Hasse diagram the restriction of \( \mathcal{Q} \) to that component. However, shrinking each cyclic component of \( \mathcal{Q} \) to a point yields a tree in which there is at most one \( d\)-path from any one vertex to another. Thus, the relation \( \preceq \) is a partial order on all of \( E \) and the Hasse diagram of this poset, called the regular poset of \( A \) and of \( R \), is \( \mathcal{Q} \). Therefore, we can assume that the labels for the elements of \( E \) have been chosen so that \( i \leq j \) if \( e_i \preceq e_j \), for all \( i, j \).

If \( \Lambda \) is an arbitrary ring, then the \( \Lambda \) incidence ring \( \Lambda E \) of the finite poset \((E, \preceq) \) can be realized as the ring of all \( n \times n \) matrices \([a_{ij}]\) over \( \Lambda \) with \( a_{ij} = 0 \) whenever \( e_i \not\preceq e_j \). If \( K \) is a field, then the \( K \)-incidence algebra \( KE \) of \((E, \preceq) \) is a subalgebra of the algebra of all \( n \times n \) upper triangular matrices over \( K \). For each \( 1 \leq i, j \leq n \) let \( e_{ij} \) be the \((i, j)\) matrix unit in \( \mathbb{M}_n(K) \). Then the \( K \)-incidence algebra \( KE \) has \( K \)-basis

\[
B(E) = \{ e_{ij} \in \mathbb{M}_n(K) \mid e_i \preceq e_j \text{ in } E \}.
\]

If \( A \) is transitive with regular poset \( E \), then in fact, \( A \) is isomorphic to the subalgebra \( K \xi E \) of upper triangular matrices with \( K \)-basis \( B(E) \) and with multiplication twisted by a two-dimensional cocycle \( \xi \) of the poset \( E \) with coefficients in \( K^* \). (See [5], [1] Theorem 2.4, or [2] Theorem 1.14.)

More generally, suppose that \( A \) is locally transitive. Consider the \( K \)-path algebra \( K \mathcal{Q} \) of the quiver \( \mathcal{Q} \) of \( A \). By Gabriel’s Theorem (see [8], Section 4; or [4] Theorem III.1.9) there is a surjective \( K \)-algebra homomorphism, called a Gabriel homomorphism for \( A \), \( \gamma : K \mathcal{Q} \rightarrow A \) from \( K \mathcal{Q} \) onto \( A \) with \( \gamma(e_i) = e_i \) for each \( e_i \in E \). Now suppose that \( p \) and \( q \) are two \( d \)-paths in \( \mathcal{Q} \) from \( e \) to \( f \) for which \( p \cup q \) forms a bi-directed cycle in \( \mathcal{Q} \). Then \( p \) and \( q \) must belong to the same cyclic component, and so since \( A \) is locally transitive,
by Lemma 4.2, $eAf \neq 0$. Thus, $\dim_K(eAf) = 1$ and so

$$eAf = K\gamma(p) = K\gamma(q).$$

Thus, for each such pair $(p, q)$ of $d$-paths in $Q$ there is a nonzero $\rho(p, q) \in K^*$ with

$$\gamma(p) = \rho(p, q)\gamma(q).$$

Next, let $I$ be the ideal

$$I = \langle p - \rho(p, q)q : p \cup q \text{ is a bidirected cycle in } Q \rangle$$

of the $K$-path algebra $KQ$. Let $\overline{A} = KQ/I$, and let $\overline{\gamma} : KQ \rightarrow \overline{A}$ be the natural surjective algebra homomorphism. Since $I \leq \ker(\gamma)$, it follows that $\gamma$ factors through $\overline{\gamma}$, say $\phi : \overline{A} \rightarrow A$ with $\gamma = \phi \circ \overline{\gamma}$. Also, if $p$ and $q$ are two $d$-paths from $e$ to $f$, then $p$ and $q$ factor as

$$p = p_1p_2\cdots p_t \quad \text{and} \quad q = q_1q_2\cdots q_t,$$

where for each $1 \leq k \leq t$, either $p_k = q_k$ or $p_k \cup q_k$ is a bi-directed cycle. Letting $\rho(p_k, q_k) = 1$ if $p_k = q_k$, we can define $\rho(p, q) \in K^*$ by

$$\rho(p, q) = \rho(p_1, q_1)\rho(p_2, q_2)\cdots \rho(p_t, q_t).$$

Then

$$\gamma(p) = \rho(p, q)\gamma(q).$$

For each $e_i \in E$, let $\overline{e_i} = e_i + I \in \overline{A}$. Note that $\{\overline{e_1}, \ldots, \overline{e_n}\}$ is a basic set of primitive idempotents for $\overline{A}$. Suppose $e, f \in E$ with $eAFf \neq 0$. A routine check shows that $\dim_K(\overline{eAFf}) = 1$ and thus, $\overline{A}$ is square-free. If $e \neq f$ and $e(KQ)f \neq 0$, then there is a path from $e$ to $f$ in $Q$. Since $I$ contains no directed paths, we can conclude that $\overline{A}$ is transitive.

Thus, $\overline{A}$ is a transitive square-free $K$-algebra and hence $\overline{A} \cong K\xi E$ for some two-dimensional cocycle $\xi$ of $E$. Combining these observations with Theorem 4.8, we have the following matrix realization of locally transitive square-free rings with solvable quivers:

**Theorem 5.1.** Let $R$ be a basic, locally transitive, indecomposable square-free ring with solvable quiver $Q$ on the basic set $E = \{e_1, \ldots, e_n\}$, and with division ring $D$. The relation defined on $E$ by

$$e_i \leq e_j \iff \text{there is a } d\text{-path in } Q \text{ from } e_i \text{ to } e_j$$

is a partial order and the Hasse diagram of the poset $(E, \leq)$ is isomorphic to the digraph $Q$. Then there is a two-dimensional cocycle $\xi$ of the poset $E$ with coefficients in $K = \text{Cen}D$ such that $R$ is isomorphic to a factor ring of the $D$ incidence ring $D\xi E = D \otimes_K (K\xi E)$ of $E$ twisted by $\xi$. Finally, $R$ is transitive iff $R$ is isomorphic to the twisted $D$-incidence ring $D\xi E$. 
Acknowledgements. The authors would like to thank Jessica K. Sklar for several helpful discussions about and comments on earlier versions of this work.

References


Received June 13, 2001 and revised April 7, 2003.

**University of Oregon**
**Eugene, OR 97403**
*E-mail address*: anderson@math.uoregon.edu

**John Carroll University**
**University Heights, OH 44118**
*E-mail address*: bdambrosia@jcu.edu
SYMMETRIES OF REAL CYCLIC $p$-GONAL RIEMANN SURFACES

ANTONIO F. COSTA AND MILAGROS IZQUIERDO

A closed Riemann surface $X$ which can be realised as a $p$-sheeted covering of the Riemann sphere is called $p$-gonal, and such a covering is called a $p$-gonal morphism. A $p$-gonal Riemann surface is called real $p$-gonal if there is an anticonformal involution (symmetry) $\sigma$ of $X$ commuting with the $p$-gonal morphism. If the $p$-gonal morphism is a cyclic regular covering the Riemann surface is called real cyclic $p$-gonal, otherwise it is called real generic $p$-gonal. The species of the symmetry $\sigma$ is the number of connected components of the fixed point set $\text{Fix}(\sigma)$ and the orientability of the Klein surface $X/\langle \sigma \rangle$. In this paper we find the species for the possible symmetries of real cyclic $p$-gonal Riemann surfaces by means of Fuchsian and NEC groups.

1. Introduction.

A closed Riemann surface $X$ which can be realised as a $p$-sheeted covering of the Riemann sphere is called $p$-gonal, and such a covering is called a $p$-gonal morphism. The $p$-gonal Riemann surfaces have been extensively studied, see \cite{[1],[2],[6],[8],[9],[12] and [13]}. A $p$-gonal Riemann surface is called real $p$-gonal if there is an anticonformal involution (symmetry) $\sigma$ of $X$ commuting with the $p$-gonal morphism.

Let $X_g$ be a real $p$-gonal Riemann surface of genus $g \geq 2$. A symmetry $\sigma$ of $X_g$ is an anticonformal involution of $X_g$. The topological type of a symmetry is determined by the number of connected components, called ovals, of the fixed-point set $\text{Fix}(\sigma)$ and the orientability of the Klein surface $X/\langle \sigma \rangle$. We say that $\sigma$ has species $\Sigma_\sigma = +k$ if $\text{Fix}(\sigma)$ consists of $k$ ovals and $X/\langle \sigma \rangle$ is orientable, and $\Sigma_\sigma = -k$ if $\text{Fix}(\sigma)$ consists of $k$ ovals and $X/\langle \sigma \rangle$ is non-orientable. The set $\text{Fix}(\sigma)$ corresponds to the real part of a complex algebraic curve representing $X$, which admits an equation with real coefficients.

If the $p$-gonal morphism is a cyclic regular covering, then the Riemann surface is called real cyclic $p$-gonal. When $p = 2$ the surface $X_g$ is called hyperelliptic. A Riemann surface represented by an algebraic curve given
by an equation of the form

$$y^p = \prod (x-a_i) \prod (x-b_j)^2 \cdots \prod (x-m_j)^{p-1}$$

where the coefficients of the polynomial $\prod (x-a_i) \cdots \prod (x-m_j)^{p-1}$ are real is a real cyclic $p$-gonal Riemann surface. The complex conjugation induces a symmetry on the above curve. A natural problem is to study and classify all possible symmetries of such a Riemann surface up to conjugacy, as they will produce non-isomorphic real models of the complex algebraic curve.

In Section 2 we characterise real cyclic $p$-gonal Riemann surfaces, where $p$ is an odd prime, in terms of signatures of Fuchsian and NEC groups. In Section 3 we determine all possible symmetries of a real cyclic $p$-gonal Riemann surface represented by an algebraic curve with equation (1.1).

2. Signatures of real cyclic $p$-gonal Riemann surfaces.

Let $X_g$ be a compact Riemann surface of genus $g \geq 2$. The surface $X_g$ can be represented as a quotient $X_g = \mathcal{H}/\Gamma$ of the upper half plane $\mathcal{H}$ under the action of a surface Fuchsian group $\Gamma$, that is, a cocompact orientation-preserving subgroup of the group $\mathcal{G} = \text{Aut}(\mathcal{H})$ of conformal and anticonformal automorphisms of $\mathcal{H}$ without elliptic elements. A discrete, cocompact subgroup $\Gamma$ of $\text{Aut}(\mathcal{H})$ is called an NEC (Non-Euclidean Crystallographic) group. The subgroup of $\Gamma$ consisting of the orientation-preserving elements is called the canonical Fuchsian subgroup of $\Gamma$, it is denoted by $\Gamma^+$. The algebraic structure of an NEC group and the geometric structure of its quotient orbifold are given by the signature of $\Gamma$:

$$s(\Gamma) = (h, \pm, [m_1, \ldots, m_r], \{(n_{11}, \ldots, n_{1s_1}), \ldots, (n_{k1}, \ldots, n_{ks_k})\}).$$

The orbit space $\mathcal{H}/\Gamma$ is an orbifold with underlying surface of genus $h$, having $r$ cone points and $k$ boundary components, each with $s_j \geq 0$ corner points. The signs $\pm$ correspond to orientable and non-orientable orbifolds respectively. The integers $m_i$ are called the proper periods of $\Gamma$ and they are the orders of the cone points of $\mathcal{H}/\Gamma$. The brackets $(n_{i1}, \ldots, n_{is_i})$ are the period cycles of $\Gamma$ and the integers $n_{ij}$ are the link periods of $\Gamma$ and the orders of the corner points of $\mathcal{H}/\Gamma$. The group $\Gamma$ is called the fundamental group of the orbifold $\mathcal{H}/\Gamma$.

A group $\Gamma$ with signature (2.1) has a canonical presentation with generators:

$$x_1, \ldots, x_r, e_1, \ldots, e_k, c_{ij}, 1 \leq i \leq k, 1 \leq j \leq s_i + 1, \quad a_1, b_1, \ldots, a_h, b_h$$

if $\mathcal{H}/\Gamma$ is orientable, or

$$d_1, \ldots, d_h$$
otherwise, and relators:

\begin{equation}
\begin{aligned}
x_{i}^{m_{i}}, & \quad i = 1, \ldots, r, \\
c_{ij}^{2}, (c_{ij-1}c_{ij})^{n_{ij}}, c_{i0}e_{i}^{-1}c_{i0}c_{i}, & \quad i = 1, \ldots, k, j = 2, \ldots, s_{i} + 1
\end{aligned}
\end{equation}

and 
\[x_{1} \cdots x_{r}e_{1} \cdots e_{k}a_{1}^{-1}b_{1}^{-1} \cdots a_{h}b_{h}a_{h}^{-1}b_{h}^{-1} \text{ or } x_{1} \cdots x_{r}e_{1} \cdots e_{k}d_{1}^{2} \cdots d_{s}^{2}\]
according to whether \(H/\Gamma\) is orientable or not. This last relation is called the long relation.

The hyperbolic area of the orbifold \(H/\Gamma\) coincides with the hyperbolic area of an arbitrary fundamental region of \(\Gamma\) and equals:

\begin{equation}
\mu(\Gamma) = 2\pi \left( \varepsilon h - 2 + k + \sum_{i=1}^{r} \left( 1 - \frac{1}{m_{i}} \right) + \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{s_{i}} \left( 1 - \frac{1}{m_{ij}} \right) \right),
\end{equation}

where \(\varepsilon = 2\) if there is a \(\prime\prime\) sign and \(\varepsilon = 1\) otherwise. If \(\Gamma'\) is a subgroup of \(\Gamma\) of finite index then \(\Gamma'\) is an NEC group and the following Riemann-Hurwitz formula holds:

\begin{equation}
[\Gamma : \Gamma'] = \mu(\Gamma') / \mu(\Gamma).
\end{equation}

An NEC group \(\Gamma\) without elliptic elements is called a surface group and it has signature \((h; \pm; [-], \{(-), k, \ldots, (-)\})\). In such a case \(H/\Gamma\) is a Klein surface, i.e., a surface with a dianalytic structure of topological genus \(h\), orientable or not according to the sign \(\prime\prime+\prime\prime\) or \(\prime\prime-\prime\prime\), and having \(k\) boundary components. Conversely, a Klein surface whose complex double has genus greater than one can be expressed as \(H/\Gamma\) for some NEC surface group \(\Gamma\). Furthermore, given a Riemann (resp. Klein) surface represented as the orbit space \(X = H/\Gamma\), with \(\Gamma\) a surface group, a finite group \(G\) is a group of automorphisms of \(X\) if and only if there exists an NEC group \(\Delta\) and an epimorphism \(\theta: \Delta \to G\) with \(\ker(\theta) = \Gamma\) (see [5]). The NEC group \(\Delta\) is the lifting of \(G\) to the universal covering \(\pi: H \to H/\Gamma\) and is called the universal covering transformation group of \((X, G)\).

**Definition 1.** For a prime \(p\), a real cyclic \(p\)-gonal Riemann surface is a triple \((X, f, \sigma)\) where \(\sigma\) is a symmetry of \(X\), \(f\) is a cyclic \(p\)-gonal morphism and \(f \circ \sigma = c \circ f\), and \(c\) is the complex conjugation.

Notice that by Lemma 2.1 in [1] the condition \(f \circ \sigma = c \circ f\) is automatically satisfied for genera \(g \geq (p - 1)2 + 1\), since the \(p\)-gonal morphism is unique. From now on, the genera will satisfy the condition above. As a consequence of the assumption \(g \geq (p - 1)2 + 1\) for the genera of the \(p\)-gonal surface \(X_{g}\) we have that the group \(C_{p}\) generated by the \(p\)-gonal morphism is a normal subgroup of \(\text{Aut}^{+}(X_{g})\). Notice the the classification method fails for surfaces with genera in the range \(2 \leq g \leq (p - 1)2\). For instance, there are two 7-gonal surfaces of genus 3. One of them, \(X_{3}\), is the Klein’s quartic with \(\text{Aut}^{+}(X_{3}) \simeq \text{PSL}_{2}(7)\), in this case \(C_{7}\) is non-normal in \(\text{PSL}_{2}(7)\).
We give now a characterisation of real cyclic $p$-gonal Riemann surfaces represented by real equations via NEC groups.

Theorem 1 ([7]). Let $X$ be a Riemann surface with genus $g$. The surface $X$ admits a symmetry $\sigma$ and a meromorphic function $f$ such that $(X, f, \sigma)$ is a real cyclic $p$-gonal Riemann surface represented by a curve with real equation $y^p = \prod (x - a_i) \cdots \prod (x - m_j)^{p-1}$ if and only if there are an NEC group $\Delta$ with signature $(0, +, \{p, \ldots, p\}, \{(p, \ldots, p)\})$ and an epimorphism $\theta : \Delta \to D_p$ such that $X$ is conformally equivalent to $\mathcal{H}/\Ker \theta$ and $\Ker \theta$ is an NEC Fuchsian surface group.

Let $(X, f, \sigma)$ be a real cyclic $p$-gonal Riemann surface uniformised by a Fuchsian surface group $\Gamma$. Consider the automorphism $\varphi : X \to X$ such that $X/\langle \varphi \rangle$ is the Riemann sphere and $\varphi$ is a deck-transformation of the covering $f$. Notice that the group $\Delta$ is the universal covering transformation group of $(X, \varphi, \sigma)$, that $D_p = \langle \varphi, \sigma \rangle$ and that the canonical Fuchsian subgroup $\Delta^+$ is the universal covering transformation group of $(X, \varphi)$. Thus $X/\langle \varphi \rangle$ is a sphere with conic points of order $p$. Let $\overline{\sigma}$ be the symmetry in the Riemann sphere $X/\langle \varphi \rangle$ induced by $\sigma$. Since the triple $(X, \varphi, \sigma)$ is represented by the equation $y^p = \prod (x - a_i) \cdots \prod (x - m_j)^{p-1}$, the symmetry $\sigma$ is given by the map $\sigma : (x, y) \to (x, \overline{y})$. The set of real solutions of 1.1 is the set Fix($\sigma$). Thus $\overline{\sigma}$ is conjugated to the complex conjugation. Then $X/\langle \varphi, \sigma \rangle = X/\langle \varphi \rangle/\langle \overline{\sigma} \rangle$ is a disc with corner(s) and conic points of order $p$.

With the above notation:

Theorem 2. Let $X$ be a real cyclic $p$-gonal Riemann surface such that $\langle \varphi, \sigma \rangle$ is isomorphic to $D_p$. If $G$ is the group of conformal and anticonformal automorphisms of $X$, then $X/G$ is uniformised by an NEC group $\Lambda$ such that there is a surface Fuchsian subgroup $\Gamma \leq \Lambda$ uniformising $X$ and the group $\Lambda$ has one of the following signatures:

(I) $(0, +, [\overline{p}, \ldots, \overline{p}], \{(\overline{p}, \ldots, \overline{p})\})$, where $\epsilon = 0$ or $1$ and $2r + s = \frac{2g + 2(1-\epsilon)(p-1)}{q(p-1)}$. $G/\langle \varphi \rangle = C_q \times C_2$.

(II) $(0, +, [\overline{p}, \ldots, \overline{p}], \{(qp^{\epsilon_1}, \overline{p}, \ldots, \overline{p}, qp^{\epsilon_2}, \overline{p}, \ldots, \overline{p})\})$, where $\epsilon_i = 0$ or $1$ and $2r + s_1 + s_2 = \frac{2g + (2-\epsilon_1-\epsilon_2)(p-1)}{q(p-1)}$. $G/\langle \varphi \rangle = D_q$.

(III) $(0, +, [\overline{p}, \ldots, \overline{p}], 2p^{\epsilon_1}, \{(qp^{\epsilon_2}, \overline{p}, \ldots, \overline{p})\})$, where $\epsilon_i = 0$ or $1$ and $2r + s = \frac{q + (1-\epsilon_1-\epsilon_2)(p-1)}{q(p-1)}$. $G/\langle \varphi \rangle = D_q \times C_2$.

(IV) $(0, +, [\overline{p}, \ldots, \overline{p}], \{(2p^{\epsilon_1}, \overline{p}, \ldots, \overline{p}, 2p^{\epsilon_2}, \overline{p}, \ldots, \overline{p}, qp^{\epsilon_3}, \overline{p}, \ldots, \overline{p})\})$, where $\epsilon_i = 0$ or $1$ and $2r + s_1 + s_2 + s_3 = \frac{2g + (2-\epsilon_1-\epsilon_2-\epsilon_3)(p-1)}{2q(p-1)}$. $G/\langle \varphi \rangle = D_q \times C_2$. 

234 A.F. COSTA AND M. IZQUIERDO
(V) \((0, +, [p, \ldots, p], \{(2p^s, p, \ldots, p, 3p^t, p, \ldots, p, 3p^t, p, \ldots, p)\})\), where \(\epsilon_i = 0\) or \(1\) and \(2r + s_1 + s_2 + s_3 = \frac{g(1-3\epsilon_1-2\epsilon_2-2\epsilon_3)(p-1)}{6(p-1)}\). 
\(G/\langle \varphi \rangle = S_4\).

(VI) \((0, +, [p, \ldots, p, 3p^s], \{(2p^t, p, \ldots, p)\})\), where \(\epsilon_i = 0\) or \(1\) and \(2r + s = \frac{g(1-4\epsilon_1-3\epsilon_2)(p-1)}{6(p-1)}\). 
\(G/\langle \varphi \rangle = A_4 \times C_2\).

(VII) \((0, +, [p, \ldots, p], \{(2p^t, p, \ldots, p, 3p^t, p, \ldots, p, 4p^t, p, \ldots, p, 5p^t, p, \ldots, p)\})\), where \(\epsilon_i = 0\) or \(1\) and \(2r + s_1 + s_2 + s_3 = \frac{g(1-15\epsilon_1-10\epsilon_2-6\epsilon_3)(p-1)}{30(p-1)}\). 
\(G/\langle \varphi \rangle = S_4 \times C_2\).

(VIII) \((0, +, [p, \ldots, p], \{(2p^t, p, \ldots, p, 3p^t, p, \ldots, p, 5p^t, p, \ldots, p)\})\), where \(\epsilon_i = 0\) or \(1\) and \(2r + s_1 + s_2 + s_3 = \frac{g(1-15\epsilon_1-10\epsilon_2-6\epsilon_3)(p-1)}{30(p-1)}\). 
\(G/\langle \varphi \rangle = A_5 \times C_2\).

Notice that in cases (VII) and (VIII) the factor group \(C_2\) of \(G/\langle \varphi \rangle\) is generated by the antipodal map.

**Proof.** Consider the chain of coverings \(X = \mathcal{H}/\Gamma \to X/\langle \varphi \rangle = \mathcal{H}/\Delta^+ \to X/G = \mathcal{H}/\Lambda\) with uniformising groups \(\Gamma \leq \Delta^+ \leq \Lambda\), where \(s(\Delta^+) = \{(0, +, [p, \ldots, p], \{\})\text{ and } s(\Lambda) = (h, \pm, [m_1, \ldots, m_r], \{(n_{11}, \ldots, n_{1s_1}), \ldots, (n_{k_1}, \ldots, n_{ks_k})\})\). Furthermore by Lemma 2.1 in [1] the group \(\langle \varphi \rangle\) is a normal subgroup of \(G\). By Theorem 1 the factor group \(\overline{G} = G/\langle \varphi \rangle\) is a finite group of conformal and anticonformal automorphisms of the Riemann sphere. (See also [12].)

In other words, we have an epimorphism \(\theta : \Lambda \to \overline{G}\) with \(\text{Ker} \theta = \Delta^+\). This yields the signature of the group \(\Lambda\) in terms of the signature of \(\Delta^+\) and the group \(\overline{G}\). Let \(p_i\) and \(q_i\) be the orders in \(\overline{G}\) of \(\theta(x_i)\) and \(\theta(c_{ij}c_{ij-1})\) respectively, where \(x_i, c_{ij}\) are generators in the canonical presentation of \(\Lambda\) associated to the signature (2.1). By [3] and [5] each proper period \(m_i\) induces \(\overline{\mathcal{C}}_{p_i}\) proper periods \(\frac{n_i}{p_i}\) in \(s(\Delta^+)\). Each link-period \(n_{ij}\) induces \(\overline{\mathcal{C}}_{q_{ij}}\) proper periods \(\frac{n_{ij}}{q_{ij}}\) in \(s(\Delta^+)\). But \(\frac{n_{ij}}{p_i} = p\) or \(\frac{n_{ij}}{p_i} = 1\) and \(\frac{n_{ij}}{q_{ij}} = p\) or \(\frac{n_{ij}}{q_{ij}} = 1\), since \(\Delta^+\) is the group of the Riemann sphere with conic points of prime order \(p\). We denote \(K_1 = \{(i, \frac{m_i}{p_i} = 1)\}, K_p = \{(i, \frac{m_i}{p_i} = p)\}, H_1 = \{(i, j), \frac{n_{ij}}{q_{ij}} = 1\}\) and \(H_p = \{(i, j), \frac{n_{ij}}{q_{ij}} = p\}\). Thus \(\rho = \sum_{i \in K_p} \overline{\mathcal{C}}_{p_i} + \sum_{(i, j) \in H_p} \overline{\mathcal{C}}_{q_{ij}}\).
Using the Riemann-Hurwitz formula $|\mathcal{G}| = \mu(\Delta^+)/\mu(\Lambda)$ we obtain

\begin{equation}
-2 + \left( \sum_{i \in K_p} \frac{|\mathcal{G}|}{p_i} + \sum_{(i,j) \in H_p} \frac{|\mathcal{G}|}{2q_{ij}} \right) \frac{(p - 1)}{p} = |\mathcal{G}|(\alpha h - 2 + k) + \sum_{i \in K_p} |\mathcal{G}| \left( 1 - \frac{1}{pp_i} \right) + \sum_{i \in K_1} |\mathcal{G}| \left( 1 - \frac{1}{p_i} \right) + \sum_{(i,j) \in H_p} \frac{|\mathcal{G}|}{2} \left( 1 - \frac{1}{pq_{ij}} \right) + \sum_{(i,j) \in H_1} \frac{|\mathcal{G}|}{2} \left( 1 - \frac{1}{q_{ij}} \right),
\end{equation}

therefore $h = 0$, $k = 1$, $s(\Lambda) = (0, +, \{pp_1, \ldots, pp_r\}, \{(pq_1, \ldots, pq_s)\})$, where $p_i, q_j \in \{1, p\}$. By setting $K_1, K_p, H_1$ and $H_p$ in Equation (2.6) we obtain that $p_i, q_j$ satisfy the equation

\begin{equation}
|\mathcal{G}| - 2 = \sum_{1}^{r} |\mathcal{G}| \left( 1 - \frac{1}{p_i} \right) + \sum_{1}^{s} \frac{|\mathcal{G}|}{2} \left( 1 - \frac{1}{q_j} \right).
\end{equation}

To find $s(\Lambda)$ it is enough to find the nontrivial solutions of (2.7). We divide the study of (2.7) in eight cases according to the factor group $\mathcal{G}$ in the epimorphism $\theta : \Lambda \to \mathcal{G}$ with $\text{Ker} (\theta) = \Delta^+$:

(I) $\mathcal{G} = C_q \times C_2$, where $C_2 = \langle \sigma \rangle$. The solution of Equation (2.7) is $p_1 = q$. Applying Riemann-Hurwitz formula to the covering $X \to X/G$ we obtain the signature $(0, +, \{p, \ldots, p, \eta \}, \{(p, \ldots, p)\})$, where $\eta = 0$ or $1$ and $2r + s = \frac{2g + 2(1 - \epsilon)(p - 1)}{q(p - 1)}$.

(II) $\mathcal{G} = D_q$. The solution of (2.7) is $q_{j_1} = q_{j_2} = q$. Therefore $s(\Lambda) = (0, +, \{p, \ldots, p\}, \{(pq_{s_1}^1, p, \ldots, p, q_{s_2}^1, p, \ldots, p)\})$, where $\epsilon_i = 0$ or $1$ and $2r + s = \frac{2g + 2(1 - \epsilon_1 - \epsilon_2)(p - 1)}{q(p - 1)}$.

(III) $\mathcal{G} = D_q \times C_2$. The solution of (2.7) is $p_1 = 2$, and $q_1 = q$. Thus $s(\Lambda)$ becomes $(0, +, \{p, \ldots, p\}, \{(pq_{s_1}^1, p, \ldots, p, 2p \eta)\}, \{(pq_{s_2}^1, p, \ldots, p)\})$, where $\epsilon_i = 0$ or $1$ and $2r + s = \frac{2g + 2(1 - \epsilon_1 - \epsilon_2)(p - 1)}{q(p - 1)}$.

(IV) $\mathcal{G} = D_q \times C_2$. The solution in this case is $q_{j_1} = q_{j_2} = 2$ and $q_{j_3} = q$. This yields $s(\Lambda) = (0, +, \{p, \ldots, p\}, \{(2p_{s_1}^1, p, \ldots, p, 2p_{s_2}^1, p, \ldots, p, q_{s_2}^1, p, \ldots, p)\}, \{(2p_{s_3}^1, p, \ldots, p, q_{s_3}^1, p, \ldots, p)\})$, where $\epsilon_i = 0$ or $1$ and $2r + s_1 + s_2 + s_3 = \frac{2g + 2(2 - \epsilon_1 - \epsilon_2)(p - 1)}{q(p - 1)}$. 


(V) $G = S_4$. The solution of (2.7) is $q_{j_1} = 2$, $q_{j_2} = q_{j_3} = 3$. Then $s(\Lambda) = (0, +, [p, \ldots, p], \{(2p^{s_1}, p, \ldots, p, 3p^{s_2}, p, \ldots, p, 3p^{s_3}, p, \ldots, p)\})$, where $\epsilon_i = 0$ or 1 and $2r + s_1 + s_2 + s_3 = \frac{g + (1 - e_1 - e_2 - e_3)(p - 1)}{6(p - 1)}$.

(VI) $G = A_4 \times C_2$. The solution of (2.7) is $p_1 = 3$, and $q_1 = 2$. Thus $s(\Lambda)$ becomes $(0, +, [p, \ldots, p], \{(2p^{s_1}, p, \ldots, p, 3p^{s_2}, p, \ldots, p)\})$, where $\epsilon_i = 0$ or 1 and $2r + s = \frac{g + (1 - e_1 - e_2 - e_3)(p - 1)}{6(p - 1)}$.

(VII) $G = S_4 \times C_2$. The solution in this case is $q_{j_1} = 2$, $q_{j_2} = 3$ and $q_{j_3} = 4$. This yields $s(\Lambda) = (0, +, [p, \ldots, p], \{(2p^{s_1}, p, \ldots, p, 3p^{s_2}, p, \ldots, p, 4p^{s_3}, p, \ldots, p)\})$, where $\epsilon_i = 0$ or 1 and $2r + s_1 + s_2 + s_3 = \frac{g + (1 - e_1 - e_2 - e_3)(p - 1)}{12(p - 1)}$.

(VIII) $G = A_5 \times C_2$. The solution now is $q_{j_1} = 2$, $q_{j_2} = 3$ and $q_{j_3} = 5$. This yields $s(\Lambda) = (0, +, [p, \ldots, p], \{(2p^{s_1}, p, \ldots, p, 3p^{s_2}, p, \ldots, p, 5p^{s_3}, p, \ldots, p)\})$, where $\epsilon_i = 0$ or 1 and $2r + s_1 + s_2 + s_3 = \frac{g + (1 - 15e_1 - 10e_2 - 6e_3)(p - 1)}{30(p - 1)}$. This finishes the proof.

3. Species of symmetries of real cyclic $p$-gonal Riemann surfaces. Let $X$ be a real cyclic $p$-gonal Riemann surface $X$ with real equation. In the next theorem we study the topological types of the possible real forms of $X$.

**Theorem 3.** Let $X$ be a real cyclic $p$-gonal Riemann surface with $p$-gonal automorphism $\varphi$ admitting a symmetry $\sigma$ with fixed points and such that $\langle \sigma, \varphi \rangle = D_p$, $p$ prime. If $\tau$ is another symmetry of $X$, then possible species of $\tau$ are (and all cases occur):

(1) $s(\Lambda)$ as in (I).
   a) $q \equiv 1 \mod (2)$. $\Sigma_\sigma = \Sigma_\tau$. If $r + \epsilon > 0$, then $\Sigma_\sigma = -1$. If $r + \epsilon = 0$, then $\Sigma_\sigma \in \{-1, +1\}$.
   b) $q \equiv 0 \mod (2)$. $\Sigma_\sigma = \Sigma_\tau$ as in case (1a) or $\Sigma_\tau = 0$.

(2) $s(\Lambda)$ as in (II).
   a) $q \equiv 1 \mod (2)$. $\Sigma_\sigma = \Sigma_\tau$ and $\Sigma_\sigma = -1$.
   b) $q \equiv 0 \mod (2)$, $q \neq 2$. $\Sigma_\sigma = -1$ and $\Sigma_\tau = -1$ or $\Sigma_\tau = +p, +1$.

(3) $s(\Lambda)$ as in (III). $\Sigma_\tau = 0$ or $\Sigma_\sigma = \Sigma_\tau$, besides $\Sigma_\sigma = -1$.

(4) $s(\Lambda)$ as in (IV).
   a) $q \equiv 1 \mod (2)$. $\{\Sigma_\sigma, \Sigma_\tau\} \subset \{\Sigma_1, \Sigma_2\}$, where $\Sigma_1 \in \{-1, +1, +p\}$. $\Sigma_2 \in \{-1, +1, +p\}$. In both cases $\Sigma_\sigma \not= +p$ and $\Sigma_\sigma \not= +1$ if $\sigma$ is of the first type.
\[ \text{b) } q \equiv 0 \mod (2), q \neq 2. \{ \Sigma_\sigma, \Sigma_\tau \} \subset \{ \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4 \}, \text{ where } \Sigma_1 = 0, \text{ and } \Sigma_2 \in \{-1, +1, +p\} \text{ and } \Sigma_1, \Sigma_3 \in \{-1, +1, +p\}. \text{ In all cases } \Sigma_\sigma \neq +p \text{ and } \Sigma_\sigma = +1 \text{ if } \sigma \text{ is of the second type.} \]

\[ q = 2. \{ \Sigma_\sigma, \Sigma_\tau \} \subset \{ \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4 \}, \text{ where } \Sigma_4 = 0, \text{ and } \Sigma_i \in \{-1, +1, +p\} \text{ for } 1 \leq i \leq 3. \text{ In all cases } \Sigma_\sigma \neq +p. \]

\[ (5) \ s(\Lambda) \text{ as in (V). } \Sigma_\tau = 0 \text{ or } \Sigma_\sigma = \Sigma_\tau, \text{ with } \Sigma_\sigma = -1. \]

\[ (6) \ s(\Lambda) \text{ as in (VI). } \Sigma_\tau = 0 \text{ or } \Sigma_\sigma = \Sigma_\tau, \text{ with } \Sigma_\sigma = -1. \]

\[ (7) \ s(\Lambda) \text{ as in (VII). } \{ \Sigma_\sigma, \Sigma_\tau \} \subset \{ \Sigma_1, \Sigma_2, 0 \}. \Sigma_1 \in \{-1, +1, +p\} \text{ and } \Sigma_2 \in \{-1, +1, +p\}, \text{ but } \Sigma_\sigma = -1. \]

\[ (8) \ s(\Lambda) \text{ as in (VIII). } \Sigma_\tau = 0 \text{ or } \Sigma_\sigma = \Sigma_\tau, \text{ with } \Sigma_\sigma = -1. \]

\[ \text{Proof.} \text{ Consider the following chain of epimorphisms: } \theta : \Lambda \to G \to \overline{G}. \text{ The symmetries of } X \text{ are symmetries } \tau \text{ in } G \text{ which are lifts of symmetries } \pi \text{ in } \overline{G}. \text{ The species of } \tau \text{ is given by the conjugacy classes of reflections in } X = \overline{\theta}^{-1}(\varphi, \tau) = \overline{\theta}^{-1}(\varphi^{-1}(\varphi)) = \theta^{-1}(\langle \tau \rangle) \text{ and the orientability of } \mathcal{H}/X. \text{ Notice that } \langle \varphi, \tau \rangle \text{ is either a cyclic group } C_2 p, \text{ or a dihedral group } D_p \text{ of order } 2p. \]

\[ \text{As in Theorem 2 we divide the proof in eight cases corresponding to the different types of groups } \overline{G} \text{ of conformal and anticonformal automorphisms of the Riemann sphere. The signature of } \Lambda \text{ in each case is given by the corresponding case in Theorem 2. } \]

\[ (1a) \overline{G} = C_q \times C_2, \ q \equiv 1 \mod (2). \text{ In this case } \overline{G} \text{ contains just one conjugacy class of symmetries and so does } G: \text{ The one represented by } \sigma. \text{ Moreover } D_p = \langle \varphi, \sigma \rangle \text{ is a normal subgroup of index } q \text{ in } G. \text{ By } [5] \text{ the signature of } \overline{\theta}^{-1}(\langle \varphi, \sigma \rangle) \text{ is } (0, +, [p, \ldots, p], \{(q), \ldots, (p)\}). \text{ By } [14] \text{ (see also } [4]) \text{ } \Sigma_\sigma = \pm 1 \text{ as } D_p = \langle \varphi, \sigma \rangle. \text{ The sign + can only occur if } \overline{\theta}^{-1}(\langle \varphi, \sigma \rangle) \text{ has no proper periods, i.e., } r + \epsilon = 0. \text{ If } s = 0, \text{ then } r + \epsilon > 0, \text{ the possible species is } -1. \]

\[ (1b) \overline{G} = C_q \times C_2 = \langle \hat{\varphi}, \hat{\sigma} \langle \hat{\rho}^q, \hat{\sigma}^2, \hat{\rho}^{-1} \hat{\sigma} \hat{\sigma} \rangle, \text{ with } q \equiv 0 \mod (2). \text{ In this case } \overline{G} \text{ contains two conjugacy classes of symmetries, with representatives namely } \hat{\sigma} \text{ and } \hat{\rho}^{q/2} \hat{\sigma} = \tau, \text{ and so does } G. \text{ To find the species of the symmetries we have to consider the normal subgroups } \overline{\theta}^{-1}(\langle \varphi, \sigma \rangle) \text{ and } \overline{\theta}^{-1}(\langle \varphi, \tau \rangle) \text{ of } \Lambda \text{ with factor group } C_q. \text{ By } [5] \text{ they have signatures } (0, +, [p, \ldots, p], \{(q), \ldots, (p)\}) \text{ and } (0, +, [p, \ldots, p], \{-\}) \text{ respectively. So species } \Sigma_\sigma \text{ is as in (1a) and } \Sigma_\tau = 0. \]

\[ (2a) \overline{G} = D_q = \langle \hat{\varphi}, \hat{\sigma} \langle \hat{\rho}^q, \hat{\sigma}^2, (\hat{\rho} \hat{\sigma})^2 \rangle, \text{ with } q \equiv 1 \mod (2). \text{ The group } \overline{G} \text{ contains one conjugacy class of symmetries and so does } G. \text{ By the epimorphism } \theta : \Lambda \to D_q \text{ the images of reflections in } \Lambda \text{ leave one fixed coset in } D_q, \text{ so} \]
we get that $\Lambda_\sigma$ has signature $(0, s_1, \ldots, p, \ldots, p), \{(\overline{p}, \ldots, \overline{p})\})$. Now, $s_1 + s_2 + r > 0$ since $\Lambda$ is a NEC group, then $\Sigma_\sigma = -1$ by [4] and [14].

(2b) $\overline{G} = D_q = \langle \overline{p}, \overline{\varphi}, \overline{\sigma}^2, (\overline{\rho} \overline{\sigma})^2 \rangle$, with $q \equiv 0 \mod (2)$. The group $\overline{G}$ (and the group $G$) contains two conjugacy classes of symmetries, with representatives namely $\overline{\sigma}$ and $\overline{\rho} \overline{\sigma} = \overline{\tau}$. To find $\Sigma_\sigma$ and $\Sigma_\tau$ we have to study the images of reflections by an epimorphism $\theta : \Lambda \to D_q$. Each of these images leaves either 2 $\overline{\sigma}$-cosets fixed and none from $\overline{\tau}$ or the other way round.

Thus the signatures of $\overline{\Lambda}_\sigma$ and $\overline{\Lambda}_\tau$ are $(0, +, [\overline{p}, \ldots, \overline{p}], \{(\overline{p}, \ldots, \overline{p})\})$ and $(0, +, [\overline{p}, \ldots, \overline{p}], \{(\overline{p}, \ldots, \overline{p})\})$. Now $\sigma$ has 1 oval and does not separate because $\overline{\theta}^{-1}(\langle \varphi, \sigma \rangle)$ contains proper periods since $s_1 + s_2 + r > 0$ and $q > 2$. If $\epsilon_1 + s_2 + \epsilon_2 > 0$, then $\langle \varphi, \tau \rangle = D_p$ and as before $\Sigma_\tau = -1$. If $\epsilon_1 + s_2 + \epsilon_2 = 0$ the signature of $\overline{\theta}^{-1}(\langle \varphi, \tau \rangle)$ becomes $(0, +, [\overline{p}, \ldots, \overline{p}], \{(-)\})$. Thus $\Sigma_\tau = -1$, if $\langle \varphi, \tau \rangle = D_p$, and $\Sigma_\sigma = +p, +1$ if $\langle \varphi, \tau \rangle = C_{2p}$.

(3) $\overline{G} = D_q \times C_2 = \langle \overline{p}, \overline{\varphi}, \overline{\sigma}_1, \overline{\sigma}_2 \mid \rho^2, \overline{\varphi}^2, \overline{\sigma}_1^2, (\overline{\sigma}_1 \overline{\sigma}_2)^q, \rho \overline{\varphi} \overline{\sigma}_1 \rho \overline{\sigma}_2 \rangle$. The group $\overline{G}$ (and $G$) contains two conjugacy classes of symmetries, with representatives namely $\overline{\sigma} = \overline{\sigma}_1$ and $\overline{\rho} \overline{\sigma} = \overline{\tau}$. The images of reflections in $\Lambda$ are all mapped to conjugate reflections in $\overline{G}$. They are conjugate to $\overline{\sigma}$ as we know that $\sigma$ has fixed points. Thus $\Sigma_\sigma = 0$. On the other hand $\overline{\Lambda}_\sigma$ has always proper periods. Therefore $\Sigma_\sigma = -1$.

(4) $\overline{G} = D_q \times C_2 = \langle \overline{\sigma}_1, \overline{\sigma}_2, \overline{\sigma}_3 \mid \overline{\sigma}_1^2, (\overline{\sigma}_1 \overline{\sigma}_2)2, (\overline{\sigma}_2 \overline{\sigma}_3)2, (\overline{\sigma}_3 \overline{\sigma}_1)^q \rangle$, with $\overline{\sigma}_2$ central in $\overline{G}$. First of all the group $\langle \varphi, \sigma_2 \rangle$ is a normal subgroup of $G$ with factor group $D_q = \langle \overline{\sigma}_1, \overline{\sigma}_3 \rangle$.

(4a) $q \equiv 1 \mod (2)$. In this case $G$ has two conjugacy classes of reflections with representatives with images $\overline{\sigma}_1$ and $\overline{\sigma}_2$. Then there are two possible species for a symmetry of $X$: $\Sigma_{\sigma_1}, \Sigma_{\sigma_2}$. The possible signatures for $\overline{\theta}^{-1}(\langle \varphi, \sigma_1 \rangle)$ are given by the epimorphism $\Lambda \to D_q$. By this epimorphism the images of $c_0$ and $c_{s_1+i}$, for $i \geq 2$, are conjugated to $\overline{\sigma}_1$ and the image of $c_1, \ldots, c_{s_1+1}$ is the identity (representing the central symmetry). Therefore $c_0, c_1, c_{s_1+1}$ and $c_{s_2+2}$ fixes $q$ $\langle \overline{\sigma}_2 \rangle$-cosets and one $\langle \overline{\sigma}_1 \rangle$-coset each, each $c_1, \ldots, c_{s_1+1}$ fixes $2q$ $\langle \overline{\sigma}_2 \rangle$-cosets (and none $\langle \overline{\sigma}_1 \rangle$-coset), and finally each $c_{s_1+i}$, $i \geq 2$ fixes two $\langle \overline{\sigma}_1 \rangle$-cosets (and none $\langle \overline{\sigma}_2 \rangle$-coset) in $\overline{G}$. Thus $\overline{\Lambda}_1$ and
\(\overline{\Lambda}_2\) have signatures
\[
(0, +, \left[ \overline{p}, \ldots, \overline{p} \right], \{( \overline{p}, \ldots, \overline{p} ) \}) \quad \text{and}
(0, +, \left[ \overline{q} \overline{q}_1 + q \overline{q}_2 + q \overline{q}_3 \right], \{( \overline{p}, \ldots, \overline{p} ) \}) \text{, see [10].}
\]

Altogether we have that \(\Sigma_1\) is \(-1\) if \(s_2 + s_3 + \epsilon_1 + \epsilon_2 + \epsilon_3 > 0\) and \(\Sigma_1\) is \(+p,+1\) if \(s_2 + s_3 + \epsilon_1 + \epsilon_2 + \epsilon_3 = 0\) and \(\langle \varphi, \sigma_1 \rangle = C_{2p}\). On the other hand \(\Sigma_2\) is \(-1\) if \(s_1 + \epsilon_1 + \epsilon_2 > 0\) and and \(r + s_2 + s_3 + \epsilon_3 > 0\), \(\Sigma_2\) is \(+1\) if \(s_1 + \epsilon_1 + \epsilon_2 > 0\) and \(r = s_2 = s_3 = \epsilon_3 = 0\), and finally \(\Sigma_2\) is \(+p,+1\) if \(s_1 + \epsilon_1 + \epsilon_2 = 0\) and \(\langle \varphi, \sigma_2 \rangle = C_{2p}\). In both cases \(\Sigma_\sigma \neq +p\) since \(\langle \varphi, \sigma \rangle = D_p\) and if \(\sigma\) is conjugate to \(\sigma_1\) then again \(\Sigma_\sigma \neq +1\). No further restrictions exist.

(4b) \(q \equiv 0 \mod (2)\). In this case \(G\) has four conjugacy classes of reflections with representatives with homomorphic images \(\overline{\sigma}_1, \overline{\sigma}_2, \overline{\sigma}_3\) and \((\overline{\sigma}_3\overline{\sigma})^q/2\overline{\sigma}_2 = \overline{\sigma}_4\). Then there are four possible species for a symmetry of \(X\): \(\Sigma_i, 1 \leq i \leq 4\). The species are also given by the epimorphism \(\Lambda \overset{\theta}{\rightarrow} D_q\). By this epimorphism the images of \(c_0\) and \(c_{s_1+s_2+i}, \) for \(i \geq 3\), are conjugated to \(\overline{\sigma}_1\), the image of \(c_1, \ldots, c_{s_1+i}\) is the identity (representing the central symmetry), and the images of \(c_{s_1+i}, \) for \(2 \leq i \leq s_2 + 2\), are conjugate to \(\overline{\sigma}_3\). First of all \(\Sigma_4 = 0\) since no images of reflections by \(\overline{\theta}\) are conjugate to \(\overline{\sigma}_4\).

If \(q = 2\), then all the 3 symmetries are central and, as in (4a) the possible species for them are \(-1,+1\) and \(+p\).

If \(q \neq 2\) then with the same procedure as in (4a) we get the following signatures for \(\overline{\Lambda}_1, \overline{\Lambda}_2\) and \(\overline{\Lambda}_3\):
\[
(0, +, \left[ \overline{p}, \ldots, \overline{p} \right], \{( \overline{p}, \ldots, \overline{p} ) \}),
(0, +, \left[ \overline{p}, \ldots, \overline{p} \right], \{( \overline{p}, \ldots, \overline{p} ) \}),
(0, +, \left[ \overline{p}, \ldots, \overline{p} \right], \{( \overline{p}, \ldots, \overline{p} ) \}).
\]

Both \(\overline{\Lambda}_1\) and \(\overline{\Lambda}_3\) must have proper periods because otherwise all parameters in the signature of \(\Lambda\) except \(\epsilon_3\) are 0 and then \(\Lambda\) is a spherical group. Therefore \(\Sigma_1\) is \(-1\) if \(s_3 + \epsilon_1 + \epsilon_3 > 0\), \(\Sigma_1\) is \(+p,+1\) if \(s_3 + \epsilon_1 + \epsilon_3 = 0\) and \(\langle \varphi, \sigma_1 \rangle = C_{2p}\). \(\Sigma_2\) is \(-1\) if \(s_1 + \epsilon_1 + \epsilon_2 > 0\) and \(r + s_2 + s_3 + \epsilon_3 > 0\), \(\Sigma_2\) is \(+1\) if \(s_1 + \epsilon_1 + \epsilon_2 > 0\), \(r = s_2 = s_3 = \epsilon_3 = 0\), and finally \(\Sigma_2\) is \(+p,+1\) if \(s_1 + \epsilon_1 + \epsilon_2 = 0\) and \(\langle \varphi, \sigma_2 \rangle = C_{2p}\). Finally \(\Sigma_3\) is \(-1\) if \(s_2 + \epsilon_2 + \epsilon_3 > 0\), \(\Sigma_3\) is \(+p,+1\) if \(s_2 + \epsilon_2 + \epsilon_3 = 0\) and \(\langle \varphi, \sigma_3 \rangle = C_{2p}\). In all cases \(\Sigma_\sigma \neq +p\) since \(\langle \varphi, \sigma \rangle = D_p\). Again \(\Sigma_\sigma \neq +1\) if \(\sigma\) is conjugate to \(\sigma_1\) or \(\sigma_3\). No further restrictions exist.
(5) and (8) $G = \langle \sigma_1, \sigma_2, \sigma_3 | \sigma_1^2, (\sigma_1 \sigma_2)^2, (\sigma_2 \sigma_3)^3, (\sigma_3 \sigma_1)^4, \rangle$, where $q = 3$ in (5) and $q = 5$ in (8). $G$, and thus $G$, contains two conjugacy classes of symmetries, with representatives namely $\sigma = \sigma_1$ and $\tau$, with $\tau$ conjugated to the antipodal map. Then $\Sigma_\sigma = \Sigma_1$ and $\Sigma_\tau = 0$. As in case (2a), by [10] and [14], given the epimorphism $\theta$, all the generating reflections of $\Lambda$ induce reflections in $\bar{\theta}^{-1}(\sigma_1)$. So $\Sigma_\sigma = -1$ as they induce also proper periods.

(6) $G = A_4 \times C_2 = \langle \rho, \sigma_1, \sigma_2 | \rho^3, \sigma_1^2, \sigma_2^2, (\sigma_1 \sigma_2)^2, \tau_1 \tau_2 \rangle$, where $C_2$ is generated by the antipodal map. With the same arguments as in (3) we obtain that $G$ has two types of symmetries with representatives $\sigma$ and $\tau$ where $\Sigma_\tau = 0$ and $\Sigma_\sigma = -1$.

(7) $G = \langle \sigma_1, \sigma_2, \sigma_3 | \sigma_1^2, (\sigma_1 \sigma_2)^2, (\sigma_2 \sigma_3)^3, (\sigma_3 \sigma_1)^4 \rangle = S_4 \times C_2$. This case is as case (4b) where the central symmetry is conjugated to the antipodal map and $\sigma_1$ is conjugated to $\sigma_2$. There are 3 conjugacy classes of symmetries with species 0, $\Sigma_1 = \Sigma_\sigma_3$ and $\Sigma_2 = \Sigma_\sigma_2$. Now $\Sigma_1$ is $-1$ if $s_3 + c_1 + c_3 > 0$, $\Sigma_1$ is $+p$, $+1$ if $s_3 + c_1 + c_3 = 0$ and $\langle \sigma, \sigma_3 \rangle$ is $C_2p$. On the other hand $\Sigma_2$ is $-1$ if $s_1 + s_2 + c_1 + c_2 + c_3 > 0$, $\Sigma_2$ is $+p$ if $s_1 + s_2 + c_1 + c_2 + c_3 = 0$ and $\langle \sigma, \sigma_2 \rangle = C_2p$. $\Sigma_\sigma \neq +p, +1$ since $\langle \sigma, \sigma \rangle = D_3$ and $\sigma$ has fixed points.

To finish we show the existence of surfaces with the desired symmetries by listing appropriate groups $G$ and epimorphisms $\theta$. The $p$-gonal surfaces with the desired symmetries will be uniformised by the groups $\text{Ker}(\theta)$. We distinguish the same eight cases as in Theorem 2.

(1) Let $G = \langle \varphi, \rho, \sigma | \varphi^i, \rho^j, \rho^k, (\varphi \sigma)^2, (\varphi \rho)^q, \rho^{-1} \sigma \rho \sigma \rangle$ and let $\theta : \Lambda \to G$ be defined by $\theta(x_i) = \varphi^{v_i}, 1 \leq i \leq r$, $\theta(x_{r+1}) = \rho \varphi^{v_{r+1}}$, $\theta(c_{2j-1}) = \sigma$, $\theta(c_{2j}) = \varphi \sigma$, $\theta(e) = \rho^{-1} \varphi^j$, where $j_1 + \cdots + j_r + c + v_{r+1} + l \equiv 0 \mod p$.

(2) Let $G = \langle \varphi, \rho, \tau, \sigma | \varphi^p, \sigma^2, \rho^2, (\varphi \rho)^q, (\sigma \tau)^2, (\varphi \tau)^2 \rangle$. Let $\theta : \Lambda \to G$ be defined by $\theta(x_i) = \varphi^{v_i}, 1 \leq i \leq r$, $\theta(x_{r+1}) = \rho \varphi^{v_{r+1}}$, $\theta(c_0) = \tau$, $\theta(c_j) = \sigma \varphi^{v_j}$ for $1 \leq j \leq s_1 + 1$, where $u_1 = c_1$ and $u_j = 1 - u_{j-1}$, $\theta(c_j) = \varphi^{v_j}$ with $s_1 + 2 \leq j \leq s_1 + s_2 + 2$ where $u_{s_1 + 2} = c_2 + 1 - u_{s_1 + 1}$ and $u_j = 1 - u_{j-1}$, $\theta(e) = \varphi^j$, where $v_1 + \cdots + v_r + l \equiv 0 \mod p$.

To obtain the species $\Sigma_\tau = +p, +1$ we consider groups $G$ with presentation $G = \langle \varphi, \rho, \sigma | \varphi^p, \sigma^2, \rho^2, (\varphi \rho)^q, (\sigma \tau)^2, (\varphi \tau)^2 \rangle$. (3) Let $G = \langle \varphi, \rho, \sigma, \tau, (\varphi \rho)^q, (\sigma \tau)^2, (\varphi \tau)^2 \rangle$. Let $\theta : \Lambda \to G$ be defined by $\theta(x_i) = \varphi^{v_i}, 1 \leq i \leq r$, $\theta(x_{r+1}) = \rho \varphi^{v_{r+1}}$, $\theta(c_0) = \tau$, $\theta(c_j) = \sigma \varphi^{v_j}$ for $1 \leq j \leq s_1 + 1$, where $u_1 = c_1$ and $u_j = 1 - u_{j-1}$, $\theta(e) = \rho \varphi^j$, where $v_1 + \cdots + v_r + l \equiv 0 \mod p$.

(4) $G = \langle \varphi, \sigma_1, \sigma_2, \sigma_3 | \sigma_1^2, (\sigma_1 \sigma_2)^2, (\sigma_2 \sigma_3)^2, (\sigma_3 \sigma_1)^q, (\varphi \sigma_1)^2 \rangle$. Let $\theta : \Lambda \to G$ be defined by $\theta(x_i) = \varphi^{v_i}, 1 \leq i \leq r$, $\theta(x_0) = \sigma_1$, $\theta(c_j) = \sigma_2 \varphi^{v_j}$ for $1 \leq j \leq s_1 + 1$, where $u_1 = c_1$ and $u_j = 1 - u_{j-1}$, $\theta(c_j) = \sigma_1 \varphi^{v_j}$, with $s_1 + 2 \leq j \leq s_1 + s_2 + 2$ where $u_{s_1 + 2} = c_2 + 1 - u_{s_1 + 1}$ and $u_j = 1 - u_{j-1}$, $\theta(e) = \sigma_1 \varphi^{v_j}$, with $s_1 + s_2 + 3 \leq j \leq s_1 + s_2 + 3$ where $u_{s_1 + s_2 + 3} =$
$\epsilon_3 + 1 - u_{s_1 + s_2 + 2}$ and $u_j = 1 - u_{j-1}$, $\theta(e) = \varphi^l$, where $v_1 + \cdots + v_r + l \equiv 0 \mod p$.

To obtain the species $+p, +1$ one or two of the relations $(\varphi \sigma_i)^2$ in the presentation of $G$ must be substituted by relations $\varphi^{-1} \sigma_i \varphi \sigma_i$.

(5) and (8) $G = \langle \varphi, \sigma_1, \sigma_2, \sigma_3 | \sigma_1^2, (\sigma_1 \sigma_2)^2, (\sigma_2 \sigma_3)^3, (\sigma_3 \sigma_1)^{q}, \varphi^p, (\varphi \sigma_i)^2 \rangle$, where $q = 3$ in (5) and $q = 5$ in (8) and let $\theta : \Lambda \to G$ be defined as in (4).

(6) $G = \langle \varphi, \rho, \sigma_1, \sigma_2 | \varphi^p, \rho^3, \sigma_1^2, (\sigma_1 \sigma_2)^2, \rho^2 \sigma_1 \rho \sigma_2, (\varphi \sigma_1)^2, (\varphi \sigma_2)^2, (\varphi \rho)^{q} \rangle$. Let $\theta : \Lambda \to G$ be defined as $\theta(x_i) = \varphi^{c_i}$, $1 \leq i \leq r$, $\theta(x_{r+1}) = \rho \varphi^{c_1 + v_1 + 1}$, $\theta(c_0) = \sigma_1$, $\theta(c_j) = \sigma_2 \varphi^{c_j}$ for $1 \leq j \leq s + 1$, where $u_1 = \epsilon_2$ and $u_j = 1 - u_{j-1}$, $\theta(e) = \rho^2 \varphi^l$, where $v_1 + \cdots + v_{r+1} + \epsilon_1 + l \equiv 0 \mod p$.

(7) $G = \langle \varphi \sigma_1, \sigma_2, \sigma_3 | \sigma_1^2, (\sigma_1 \sigma_2)^2, (\sigma_2 \sigma_3)^3, (\sigma_3 \sigma_1)^4, \varphi^p, (\varphi \sigma_i)^2 \rangle$ and let $\theta : \Lambda \to G$ be defined as in (4). To obtain the species $+p, +1$ either the relations $(\varphi \sigma_1)^2$ and $(\varphi \sigma_3)^2$ or the relation $(\varphi \sigma_2)^2$ in the presentation of $G$ must be changed to the corresponding commuting relation.

The kernels of the above epimorphisms will uniformise surfaces with a symmetry with species $-1$ for general groups $\Lambda$. The same epimorphisms yield the species $+1$ in cases 1 and 4 under the restrictions on $\Lambda$ given in the first part of the theorem. Again, the same epimorphisms yield the species $+p, +1$ under the corresponding restrictions on $\Lambda$ given in the first part of the theorem.

The authors wish to thank to the referee for several helpful suggestions.

References


Received May 21, 2002 and revised May 8, 2003. The first author is partially supported by Ministerio de Ciencia y Tecnologia BFM2002-04801. The second author thanks the hospitality of UNED.

DEPTO. MATEMÁTICAS FUNDAMENTALES
UNED, C/ SENDA DEL REY 9
28040 MADRID
SPAIN
E-mail address: acosta@mat.uned.es

MATHEMATiska INSTITUTIONEN
LINKÖPINGS UNIVERSITET
581 83 LINKÖPING
SWEDEN
E-mail address: miizq@mai.liu.se
THE FINITE SIMPLE GROUPS HAVE COMPLEMENTED
SUBGROUP LATTICES

MAURO COSTANTINI AND GIOVANNI ZACHER

We prove that the lattice of subgroups of every finite simple
group is a complemented lattice.

1. Introduction.

A group $G$ is called a $K$-group (a complemented group) if its subgroup lattice
is a complemented lattice, i.e., for a given $H \leq G$ there exists a $X \leq G$ such
that $\langle H, X \rangle = G$ and $H \wedge X = 1$. The main purpose of this Note is to answer
a long-standing open question in finite group theory, by proving that:

Every finite simple group is a $K$-group.

In this context, it was known that the alternating groups, the projective
special linear groups and the Suzuki groups are $K$-groups ([P]).

Our proof relies on the FSGC-theorem and on structural properties of the
maximal subgroups in finite simple groups. The rest of this paper is divided
into four sections. In Section 2 we collect some criteria for a subgroup of
a group $G$ to have a complement and recall some useful known results. In
Section 3 we deal with the classical groups, in 4 with the exceptional groups
of Lie type and in Section 5 with the sporadic groups.

With reference to notation and terminology, we shall follow closely those
in use in [P] and [S]. All groups are meant to be finite.

2. Preliminaries.

We begin with the following:

Proposition 2.1. Given the group $G$, let $T, X$ be subgroups of $G$ such that
$T \leq X < G$. If the interval $[X/T]$ is a complemented lattice and if $X$
is contained in only one maximal subgroup $M$ of $G$, then every $H \leq G$ with
$H \not\leq M$ and $H \wedge T = 1$ has a complement in $G$.

Proof. Let $C$ be a complement of $\langle H, T \rangle \wedge X$ in $[X/T]$. Then $\langle H, C \rangle = \langle H, T, C \rangle \geq \langle \langle H, T \rangle \wedge X, C \rangle = X$. Since $H \not\leq M$, we conclude that $\langle H, C \rangle = G$. Moreover $H \wedge C = H \wedge X \wedge C \leq \langle H, T \rangle \wedge X \wedge C = T$, hence $H \wedge C \leq H \wedge T = 1$. \qed
The condition on $M$ in Proposition 2.1 means that $[G/X]$ is a mono-coatomic interval with coatom $M$.

**Corollary 2.2.** Let $X$ be a $K$-subgroup and $[G/X]$ a monocoatomic interval with coatom $M$. Then every $H \leq G$ not contained in $M_G$ has a complement in $G$. In particular $G/M_G$ is a $K$-group.

**Proof.** There exists a $g \in G$ such that $H^g \not\leq M$. By Proposition 2.1 with $T = 1$, $H^g$ has a complement. Hence also $H$ has a complement $C$ in $G$. Moreover, if $M_G < H$, then $CM_G/M_G$ is a complement of $H/M_G$ in $G/M_G$. □

**Proposition 2.3.** Let $G$ be a simple group and $[G/X]$ a monocoatomic interval with coatom $M$. If $N$ is a central subgroup of $M$ of prime order with $N \leq X$ and if $X/N$ is a $K$-group, then $G$ is a $K$-group.

**Proof.** Let $H$ be a proper subgroup of $G$. Since $M_G = 1$, without loss of generality we may assume $H \not\leq M$. If now $H \cap N = 1$, by Proposition 2.1 $H$ has a complement in $G$. Assume now $N \leq H$; there exists a $g \in G$ such that $N^g \not\leq H$. So if $H$ has no complement in $G$, by Proposition 2.1 we must have $N^g \leq C(H)$. It follows that if $F = \{N^x \mid x \in G\}$ and $F_1 = \{N^x \mid N^x \not\leq H\}$, then $N(H) \geq \langle H, F_1 \rangle \geq \langle F \rangle = G$, a contradiction. □

We finally recall:

(2.1) *The direct product of a family of groups is a $K$-group if and only if each factor is a $K$-group,*

see Corollary 3.1.5 in [S].

(2.2) *If $G$ contains an abelian subgroup $A$ generated by minimal normal subgroups of $G$ and a complement $K$ to $A$ that is a $K$-group, then $G$ is a $K$-group,*

see Lemma 3.1.9 in [S].

(2.3) *The symmetric and alternating groups, the projective special linear groups $L_n(q)$ and the simple Suzuki groups $^2B_2(q)$ are $K$-groups,*

see [P].

For our purpose it will be convenient to know which non-simple groups of Lie type ([C], p. 175, p. 268) are complemented.

**Proposition 2.4.** The following non-simple groups of Lie type are $K$-groups:

$L_2(2)$, $L_2(3)$, $Sp_4(2)$, $G_2(2)$, $^2G_2(3)$.

The following non-simple groups of Lie type are not $K$-groups:

$^2B_2(2)$, $^2F_4(2)$, $U_3(2)$.

**Proof.** In fact $L_2(2) \cong S_3$, $L_2(3) \cong A_4$, $Sp_4(2) \cong S_6$, and we are done by (2.3). In $G_2(2)$ there is a monocoatomic interval $[G_2(2)/H]$ with $H \cong L_3(2)$ and corefree coatom, by Theorem 2.5 in [Co]: Hence $G_2(2)$ is a $K$-group by
The group $^2G_2(3)$ has a corefree maximal subgroup isomorphic to $Z_7: Z_6$ ([K3]): Hence it is a $K$-group by (2.2). On the other hand, we have $^2B_2(2) ≅ Z_5 : Z_4$ ([A]), $U_3(2) ≅ 3^2 : Q_8$ ([KL], p. 43) and finally $[^2F_4(2) : (^2F_4(2)')_L] = 2$, but all involutions of $^2F_4(2)$ are contained in $^2F_4(2)'$ ([AS], p. 75).

To prove the main theorem, we take a counterexample $L$ of minimal order and show that such a group $L$ does not exist.

3. The simple classical groups.

We are going to assume in this section that $L = G_0(n, q)$, a (simple) classical group as in [KL].

a) $G_0(n, q)$ is not of type $A_m$, $n = m + 1$, $m ≥ 1$.

See (2.3).

b) $G_0(n, q)$ is not of type $C_m$, $n = 2m$, $m ≥ 2$.

Proof. Let $r$ be a prime divisor of $m$, so that $m = rt$, $t ≥ 1$. By Theorem 1 and Theorem 2 in [L], the interval $[PSp(2m, q) / PSp(2t, q^r)]$ is mono-coatomic. Moreover $PSp(2t, q^r)$ is simple, since $q^r ≥ 4$, of order less than the order of $L$, hence a $K$-group. But then by Corollary 2.2, $L$ is a $K$-group, a contradiction. □

c) $G_0(n, q)$ is not of type $^2A_m$, $n = m + 1$, $m ≥ 2$.

Proof. We consider first the cases $(n, q) = (3, 3), (3, 5)$. The groups $U_3(3)$ and $U_3(5)$ are $K$-groups: In fact one has $PSL_2(7) < U_3(3)$ and $A_7 < U_3(5)$ ([K1], §5). Assume now $(n, q) ≠ (3, 3), (3, 5)$. With reference to the notation in [BGL], p. 388, let $G$ be the simple adjoint algebraic group over $\mathbb{F}_q$ with associated Dynkin diagram of type $A_m$, $λ = σ_q$ and $μ = 2σ_q$: We have $G_λ = PGL_n(q)$, $G_μ = PGU_n(q)$, $O^p(G_λ) = L_n(q)$, $O^p(G_μ) = U_n(q) = G_0(n, q)$,

$$T := O^p(G_μ ∩ G_λ) = \begin{cases} PSp_n(q) & \text{if } n \text{ is even} \\ Ω_n(q) & \text{if } nq \text{ is odd} \\ Sp_{n-1}(q) & \text{if } n \text{ is odd and } q \text{ is even.} \end{cases}$$

From Theorem 2 in [BGL] it follows that $[U_n(q) / T]$ is mono-coatomic. Moreover, $T$ is a $K$-group, either because it is simple of order less than $|L|$, or because it is isomorphic to $Sp_4(2)$ (Proposition 2.4): Hence $G_0(n, q)$ is a $K$-group, a contradiction. □

d) $G_0(n, q)$ is not of type $B_m$, $n = 2m + 1$, $m ≥ 3$, $q$ odd.
Proof. Assume \( q = p^f \), with \( f > 1 \) and let \( r \) be a prime divisor of \( f \). Then by Theorem 1 in [BGL], \([PO_n(q)/PO_n(q^{1/r})]\) is monoatomic, a contradiction. Therefore we must have \( q = p \). Now, by §5 in [K1] and Proposition 4.2.15 in [KL], \( G_0(n,q) \) contains a maximal subgroup \( M \) which is a split extension of an irreducible elementary abelian 2-group by \( A_n \) or \( S_n \). Therefore \( M \) is a \( K \)-group by (2.2), and \( G_0(n,q) \) is a \( K \)-group, a contradiction. □

e) \( G_0(n,q) \) is not of type \( D_m \), \( n = 2m, m \geq 4 \).

Proof. Let \( V = \mathbb{F}_q^n \) be the natural (projective) module for \( G_0(n,q) \), and let \( W \) be a nonsingular subspace of \( V \) of dimension 1. Since \( \Omega := G_0(n,q) \) is a counterexample of minimal order, the socle \( \text{soc } H_\Omega \) of the stabilizer \( H_\Omega \) of \( W \) in \( \Omega \), which is isomorphic to \( \Omega_{n-1}(q) \) if \( q \) is odd, and to \( Sp_{n-2}(q) \) if \( q \) is even, must be contained, by Corollary 2.2, in an element \( K_\Omega \) of \( \mathcal{C}(\Omega) \cup S \) different from \( H_\Omega \) (for the definition of the family \( \mathcal{C}(\Omega) \cup S \) we refer to §1.1 and §3.1 in [KL]).

By order considerations, one can prove that only condition (i) of Theorem 4.2 in [Li] applies: This means that \( K_\Omega \) must be an element of \( \mathcal{C}(\Omega) \). Since \( H_\Omega \in \mathcal{C}_1 \), one is left to show that there does not exist an element \( K_\Omega \) in \( \mathcal{C}_i \), for an \( i \neq 1 \), such that \( \text{soc } H_\Omega < K_\Omega < \Omega \).

For \( q \) odd, the arguments used in the proof of Proposition 7.1.3 in [KL] show that such a \( K_\Omega \) does not exist, taking into account that in our situation \( n_2 = n-1 \geq 7 \). To deal with the case when \( q \) is even, again one can proceed using arguments suggested in the proof of Lemma 7.1.4 in [KL]. □

f) \( G_0(n,q) \) is not of type \( ^2D_m \), \( n = 2m, m \geq 4 \).

Proof. Following the notation in [BGL], let \( G \) be the simple adjoint algebraic group over \( \overline{\mathbb{F}}_q \) with associated Dynkin diagram of type \( D_m \), \( \lambda = \sigma_q \) and \( \mu = ^2\sigma_q \). Then \( O^\sigma(G_\lambda) = PO_n^+(q), O^\mu(G_\mu) = PO_n^-(q) = G_0(n,q), \)

\[
T := O^\mu(G_\mu \cap G_\lambda) = \begin{cases} 
\Omega_{n-1}(q) & \text{if } q \text{ is odd} \\
Sp_{n-2}(q) & \text{if } q \text{ is even}.
\end{cases}
\]

By Theorem 2 in [BGL], \([G_0(n,q)/T]\) is monoatomic. Since \( n \geq 8 \), \( T \) is simple, hence \( G_0(n,q) \) is a \( K \)-group, a contradiction. □

We have therefore completed the proof that \( L \) is not a classical group.

4. The simple exceptional groups of Lie type.

Now we are going to show that the minimal counterexample \( L \) cannot be an exceptional group of Lie type \( G(q) \).

a) \( G(q) \) is not of type \( G_2, ^2G_2 \).
Proof. If $r$ is a prime divisor of $f$, where $q = p^r$, write $q = q_0^r$. Then $G(q_0) < G(q)$ ([Co], Theorem 2.3, 2.4, [K3], Theorem A, C). Hence by Proposition 2.4, we have $L = G_2(p)$, for an odd prime $p$. But then $G_2(2)$ is maximal in $G_2(p)$ by [K3], and we are done by Proposition 2.4.

b) $G(q)$ is not of type $F_4$.

Proof. $F_4(q)$ contains a quasisimple maximal subgroup $M$ of type $B_4(q)$, with $|Z(M)| = (2, q - 1)$ ([LS], p. 322). But then, by Proposition 2.3, $F_4(q)$ is a $K$-group.

c) $G(q)$ is not of type $E_6, E_7, E_8$.

Proof. We have $F_4(q) < E_6(q)$ ([LS], Table 1), which excludes $E_6$.

If $L$ is of type $E_7$, there exist subgroups $H < M < G$ such that $|M : H| = |Z(H)| = (2, q - 1)$ and $H/Z(H) \cong L_2(q) \times P\Omega^+_2(2)$ ([LS], Table 1). Hence $H/Z(H)$ is a $K$-group by (2.1). We claim that $[G/H]$ is monocoatomic. Clear if $q$ is even. For $q$ odd, suppose $H < M_1 < G$, with $M_1 \neq M$. Since $|M : H| = 2$, we have $|M_1| > |M| \geq q^{64}$. By the Theorem in [LS], $M_1$ either is a parabolic subgroup, or it appears in Table 1 in [LS]: However, both situations are excluded by rank or order considerations. So again by Proposition 2.3, $G$ is a $K$-group, a contradiction.

Finally assume $G$ is of type $E_8$. There exist subgroups $H < M < G$ such that $|M : H| = |Z(H)| = (2, q - 1)$, with $H/Z(H) \cong P\Omega^+_4(2)$ ([I], p. 286, [LS], Table 1), hence a $K$-group. Using the Theorem in [LS] again one shows that $[G/H]$ is monocoatomic, giving rise to a contradiction.

d) $G(q)$ is not of type $2B_2$.

See (2.3).

e) $G(q)$ is not of type $2F_4$.

Proof. The group $2F_4(2)$ is not simple, and we have seen that it is not a $K$-group (Proposition 2.4). Its derived subgroup (the Tits group) is simple and it is a $K$-group, since it has a maximal subgroup isomorphic to $L_2(25)$ ([A]). So now assume $L = 2F_4(2^{2m+1})$, with $m \geq 1$. By the Main Theorem in [M], there exist $H < M < L$ such that $|M : H| = 2$ and $H \cong Sp_4(2^{2m+1})$. Since the nonabelian composition factors of maximal subgroups of $L$ not conjugate to $M$ are of type $A_1(q), 2B_2(q), U_3(q)$ and $2F_4(q^{1/r})$, $r$ an odd prime, one concludes that $[G/H]$ is monocoatomic.

f) $G(q)$ is not of type $2E_6$.

Proof. In fact we have $F_4(q) < 2E_6(q)$ from Table 1 in [LS].
g) \( G(q) \) is not of type \(^3D_4\).

\textit{Proof.} From the Theorem in [K2], we have \( G_2(q) < ^3D_4(q) \). Since \( G_2(q) \) is a \( K \)-group, we get a contradiction. \( \square \)

This concludes the proof that \( L \) is not a group of Lie type.

### 5. Sporadic simple groups.

We are left to deal with the sporadic groups: To this end, for each group we exhibit a maximal subgroup which is a \( K \)-group. From the tables in [A] we have:

\[
\begin{align*}
L_2(11) &\leq M_{11}, \quad L_2(11) \leq M_{12}, \quad A_7 \leq M_{22}, \quad M_{22} \leq M_{23}, \quad M_{23} \leq M_{24}, \\
L_2(11) &\leq J_1, \quad A_5 \leq J_2, \quad L_2(19) \leq J_3, \quad 43 : 14 \leq J_4, \quad M_{22} \leq HS, \\
A_7 \leq Suz, \quad M_{22} \leq McL, \quad A_8 \leq Ru, \quad S_4 \times L_3(2) \leq He, \quad 67 : 22 \leq L_9, \\
A_7 \leq O'N, \quad M_{23} \leq Co_2, \quad M_{23} \leq Co_3, \quad Co_3 \leq Co_1, \quad S_{10} \leq Fi_{22}, \\
S_{12} \leq Fi_{23}, \quad Fi_{23} \leq Fi'_{24}, \quad A_{12} \leq HN, \quad S_5 \leq Th, \quad 31 : 15 \leq BM, \\
31 : 15 \times S_3 \leq M.
\end{align*}
\]

We have thus completed the proof of the main theorem:

\textbf{Theorem.} Every finite simple group is a \( K \)-group.

\textbf{Acknowledgements.} We are grateful to E. Vdovin for helpful discussions.

\textbf{References}


FINITE SIMPLE GROUPS ARE $K$-GROUPS


[K3] ______, The maximal subgroups of the Chevalley groups $G_2(q)$ with $q$ odd, the Ree groups $^2G_2(q)$, and their automorphism groups, J. Algebra, 117 (1988), 30-71, MR 0955589, Zbl 0651.20020.


Received December 24, 2002 and revised February 24, 2003.

DIPARTIMENTO DI MATEMATICA PURA ED APPLICATA
UNIVERSITÀ DI PADova
VIA BELZONI 7, 35131 PADova
ITALY
E-mail address: costantini@math.unipd.it

DIPARTIMENTO DI MATEMATICA PURA ED APPLICATA
UNIVERSITÀ DI PADova
VIA BELZONI 7, 35131 PADova
ITALY
E-mail address: zacher@math.unipd.it
We provide a rigorous mathematical foundation to the study of strongly rational, holomorphic vertex operator algebras $V$ of central charge $c = 8, 16$ and 24 initiated by Schellekens. If $c = 8$ or 16 we show that $V$ is isomorphic to a lattice theory corresponding to a rank $c$ even, self-dual lattice. If $c = 24$ we prove, among other things, that either $V$ is isomorphic to a lattice theory corresponding to a Niemeier lattice or the Leech lattice, or else the Lie algebra on the weight one subspace $V_1$ is semisimple (possibly 0) of Lie rank less than 24.

1. Introduction.

One of the highlights of discrete mathematics is the classification of the positive-definite, even, unimodular lattices $L$ of rank at most 24, due originally to Minkowski, Witt and Niemeier (cf. [CS] for more information and an extensive list of references). One knows (loc. cit.) that such a lattice has rank divisible by 8; that the $E_8$ root lattice is (up to isometry) the unique example of rank 8; that the two lattices of type $E_8 + E_8$ and $\Gamma_16$ are the unique examples of rank 16; and that there are 24 inequivalent such lattices of rank 24. In each case, the lattice may be characterized by the nature of the semi-simple root system naturally carried by the set of minimal vectors (i.e., those of squared length 2).

The theory of vertex operator algebras is a newer subject which enjoys several parallels with lattice theory. Already in [B], Borcherds pointed out that one can naturally associate a vertex operator algebra $V_L$ to any positive-definite, even lattice $L$ (cf. [FLM] for a complete discussion). It is known that $V_L$ is rational and that it is holomorphic precisely when $L$ is self-dual ([D] and [DLM1]) (we defer the formal introduction of technical definitions concerning vertex operator algebras until Section 2). Since the central charge $c$ of the vertex operator algebra $V_L$ is precisely the rank of $L$, the classification of holomorphic vertex operator algebras of central charge at most 24 may be construed as a generalization of the corresponding problem for even, unimodular lattices. We will say that a holomorphic vertex operator algebra has small central charge in case it satisfies $c \leq 24$. 

253
Schellekens was the first to consider the problem of classifying holomorphic vertex operator algebras \( V \) of small central charge [\textit{Sch}]. Based on extensive computation, Schellekens wrote down a list of 71 integral \( q \)-expansions \( q^{-1} + \text{constant} + 196884q + \cdots \) and, among other things, conjectured that the graded character of a \( V \) satisfying \( c = 24 \) is necessarily equal to one of these 71 \( q \)-expansions. As is well-known, it is only the constant term that distinguishes the \( q \)-expansions from each other, the constant in question being the dimension of the Lie algebra naturally defined on the weight one subspace \( V_1 \) of \( V \) (see below for more details). Schellekens in fact wrote down a list of 71 Lie algebras (including levels) which are the candidates for \( V_1 \). (It turns out that if \( V \) has central charge strictly less that 24 then the graded character of \( V \) is uniquely determined, so at least as far as the dimension of the weight one subspace is concerned, the case \( c = 24 \) carries the most interest.)

The purpose of the present paper is to put the Schellekens program on a firm mathematical foundation within the general context of rational conformal field theory, and to make a start towards the classification (up to isomorphism) of the holomorphic vertex operator algebras of small central charge. In addition to unitarity, there are other (unstated) assumptions in [\textit{Sch}]. This circumstance means that we cannot assume the results of (loc. cit.) and need to find new approaches. In some ways we will go much further than [\textit{Sch}] in that we will be able to give an adequate characterization of the holomorphic lattice theories among all holomorphic theories of small central charge. On the other hand, although we are able to establish numerical restrictions on the nature of the Lie algebra on \( V_1 \) which show that there are only a finite number of possibilities, we are at present unable to show, assuming \( c = 24 \), that it is necessarily one of the 71 on Schellekens’ list. The remaining obstacle is essentially that of establishing\(^1\) that the levels of the associated Kac-Moody Lie algebras are positive integers (which is immediate if unitarity is assumed).

We now state some of our main results, and for this purpose we will take \( V \) to be a holomorphic vertex operator algebra of CFT type (see Section 2) which is \( C_2 \)-cofinite. (In the language of [\textit{DM}], \( V \) is strongly rational and holomorphic.) By results of Zhu [\textit{Z}], this implies that \( c \) is a positive integer divisible by 8.

**Theorem 1.** Suppose that \( c = 8 \). Then \( V \) is isomorphic to the lattice theory \( V_{E_8} \) associated to the \( E_8 \) root lattice.

**Theorem 2.** Suppose that \( c = 16 \). Then \( V \) is isomorphic to a lattice theory \( V_L \) where \( L \) is one of the two unimodular rank 16 lattices.

\(^1\)However, see comments added in proof at the end of the paper.
Thus for the cases $c = 8$ and $16$, the classification of the holomorphic vertex operator algebras completely mirrors that of the corresponding lattices. These two theorems are commonly assumed in the physics literature, and are related to the uniqueness of the heterotic string ([Sch] and [GSW]).

**Theorem 3.** Suppose that $c = 24$. Then the Lie algebra on $V_1$ is reductive, and exactly one of the following holds:

(a) $V_1 = 0$.
(b) $V_1$ is abelian of rank $24$. In this case $V$ is isomorphic to the lattice theory $V_\Lambda$ where $\Lambda$ is the Leech lattice.
(c) $V_1$ is a semi-simple Lie algebra of rank $24$. In this case $V$ is isomorphic to the lattice theory $V_L$ where $L$ is the even, unimodular rank $24$ lattice whose root system is the same as the root system associated to $V_1$.
(d) $V_1$ is a semi-simple Lie algebra of rank less than $24$.

We actually establish more than is stated here. It is a well-known conjecture [FLM] that the Moonshine module is characterized among all $c = 24$ holomorphic theories by the condition $V_1 = 0$. Our methods are less effective when there is no Lie algebra, however we will show that in this case $V_2$ carries the structure of a simple, commutative algebra (of dimension $196,884$). The commutativity and dimension formula are well-known; it is the simplicity that is novel here. (The inherent difficulty in dealing with $V_2$ is that it is not an associative algebra, indeed it is not even power associative, and there seem to be no useful identities which are satisfied.) In Case (d) we show that the simple components $g_i$ of $V_1$ have levels $k_i$ and dual Coxeter numbers $h^\vee_i$ such that the identity

$$h^\vee_i \frac{k_i}{k_i} = \frac{\dim V_1 - 24}{24}$$

holds for each $g_i$. This implies that there are only finitely many choices for the family of pairs $(g_i, k_i)$ determined by $V_1$. Note that the condition (1.1) was already identified by Schellekens [Sch]. We are in fact able to extract some further numerical restrictions on the levels $k_i$, but these fall well short of the expectation that they are all positive integers, and we forgo any discussion of this beyond (1.1). We also establish that if $V_1$ is semisimple in Theorem 3, then the Virasoro element in $V$ coincides with the usual Virasoro element associated with an affine Lie algebra defined by the Sugawara construction. We expect the result to be useful in further analysis of the situation.

The main inspiration for the proof of our results originates from our recent paper [DM]. There, we introduced methods based on the theory of modular forms used in tandem with techniques from vertex operator algebra theory. In particular we obtained a simple numerical characterization of the lattice vertex operator algebras among all rational vertex operator
algebras. The present paper is in many ways a continuation and elaboration of [loc.cit.]. When the central charge is small, the theory of modular forms gives very precise numerical information about the vertex operator algebra which allows us to show, under certain circumstances, that the numerical characterizations of [DM] are applicable. This leads to Theorems 1-3. Modular-invariance also underlies the other results that we obtain.

In addition to the Schellekens program that we have already discussed, another potential application of Theorem 3 is to the FLM conjecture regarding the Moonshine module alluded to above. Namely, Theorem 3 shows that the Leech lattice theory $V_\Lambda$ is the only $c = 24$ holomorphic theory for which the Lie algebra $V_1$ is both nonzero and not semi-simple. On the other hand, the Moonshine module is closely related to $V_\Lambda$, being built from it by a $\mathbb{Z}_2$-orbifold construction [FLM].

The paper is organized as follows: We gather together some preliminaries in Section 2, and prove Theorems 1-3 together with the supplementary result in Case (a) of Theorem 3 in Section 3. In Section 4 we identify the Virasoro element with the Sugawara construction.

2. Preliminary results.

For general background on the theory of vertex operator algebras, we refer the reader to [FLM] and [FHL]. As usual, for a state $v$ in a vertex operator algebra $V$, we denote the corresponding vertex operator by

$$Y(v, z) = \sum_{n \in \mathbb{Z}} v_n z^{-n-1},$$

while the vertex operator corresponding to the conformal (Virasoro) vector is

$$Y(\omega, z) = \sum_{n \in \mathbb{Z}} L(n) z^{-n-2}.$$

$V$ is called rational if all admissible $V$-modules are completely reducible (cf. [DLM1] for the definition of admissible module). It was shown in [DLM2] that this implies that $V$ has only finitely many inequivalent simple modules. A rational vertex operator algebra is called holomorphic if it has a unique simple module, namely the adjoint module $V$. $V$ is said to be $C_2$-cofinite in case the subspace spanned by elements $u_{-2}v$ for $u, v \in V$ is of finite codimension. Finally, we say that $V$ is of CFT-type in case the natural $\mathbb{Z}$-grading on $V$ takes the form

$$V = V_0 \oplus V_1 \oplus \cdots$$

with $V_0 = \mathbb{C}1$.

Throughout the rest of this paper, we assume that $V$ is a $C_2$-cofinite, holomorphic vertex operator algebra of CFT-type of small central charge $c \leq 24$. In order to avoid the case of the trivial vertex operator algebra $V = \mathbb{C}$, we also assume that $V$ has dimension greater than one.
Next we discuss some consequences of these assumptions for the structure of \( V \). Many of them are well-known. First, the weight 1 subspace \( V_1 \) of \( V \) carries a natural structure of Lie algebra given by

\[
[u, v] = u_0 v
\]

for \( u, v \in V_1 \). Because the adjoint module is the unique simple \( V \)-module, then the contragredient module \( V' \) is necessarily isomorphic to \( V \). This is equivalent to the existence of a nondegenerate, invariant bilinear pairing

\[
\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}
\]

which is necessarily symmetric. (For the theory of contragredient modules, cf. Section 3 of [FHL].) Because of (2.1), Li’s theory of invariant bilinear forms [L1] shows that we have \( L(1) V_1 = 0 \) and that \( \langle \cdot, \cdot \rangle \) is uniquely determined up to an overall scalar. It is convenient to fix the normalization so that

\[
\langle 1, 1 \rangle = -1.
\]

In particular, the restriction of \( \langle \cdot, \cdot \rangle \) to \( V_1 \) is given by

\[
\langle u, v \rangle |_{V_1} = u_1 v
\]

for \( u, v \in V_1 \).

In the language of [DM], it follows from what we have said that \( V \) is strongly rational, so that the results of (loc. cit.) apply. They tell us that the following hold:

(I) \( V_1 \) is a reductive Lie algebra of Lie rank \( l \leq c \).

(II) \( l = c \) if, and only if, \( V \) is isomorphic to a lattice theory \( V_L \) for some positive-definite, even, unimodular lattice \( L \).

We also note that by results of Zhu [Z] (cf. [DLM3]), \( c \) is necessarily a positive integer divisible by 8. So in fact \( c = 8, 16, \) or 24.

We refer the reader to [Z] and [DM] for an extended discussion of the role of modular-invariance in the theory of rational vertex operator algebras.

We need to recall the vertex operator algebra \( (V, Y(\cdot), 1, \omega - c/24) \) from [Z]. The new vertex operator associated to a homogeneous element \( a \) is given by

\[
Y[a, z] = \sum_{n \in \mathbb{Z}} a[n] z^{-n-1} = Y(a, e^z - 1) e^{z \omega a}
\]

while the Virasoro element is \( \tilde{\omega} = \omega - c/24 \). Thus

\[
a[m] = \text{Res}_z \left( Y(a, z)(\ln(1 + z))^m(1 + z)^{wta - 1} \right)
\]

and

\[
a[m] = \sum_{i=m}^{\infty} c(wta, i, m) a(i)
\]
for some scalars \( c(wta, i, m) \) such that \( c(wta, m, m) = 1 \). In particular,

\[
a[0] = \sum_{i \geq 0} \binom{wta - 1}{i} a(i).
\]

We also write

\[
Y[w - c/24, z] = \sum_{n \in \mathbb{Z}} L[n] z^{-n-2}.
\]

Then the \( L[n] \) again generate a copy of the Virasoro algebra with the same central charge \( c \).

Now \( V \) is graded by the \( L[0] \)-eigenvalues, that is

\[
V = V[0] \oplus V[1] \oplus \cdots
\]

where \( V[n] = \{ v \in V | L[0]v = nv \} \).

We will need the following facts: Let \( v \in V \) satisfy \( L[0]v = kv \). Then the graded trace

\[
Z(v, \tau) = \text{Tr}_V o(v) q^{L(0) - c/24} = q^{-c/24} \sum_{n \geq 0} \text{Tr}_{V_n} o(v) q^n
\]

is a modular form on \( \text{SL}(2, \mathbb{Z}) \) (possibly with character), holomorphic in the complex upper half-plane \( H \) and of weight \( k \). Here, we have set \( o(v) \) to be the zero mode of \( v \), defined via \( o(v) = v_{wt-1} \) if \( v \) is homogeneous, and extended by linearity to the whole of \( V \). Moreover, \( \tau \) will denote an element in \( H \) and \( q = e^{2\pi i \tau} \). In particular, if we take \( v \) to be the vacuum element then (2.5) is just the graded trace

\[
\text{ch}_q V = q^{-c/24} \sum_{n \geq 0} (\text{dim} V_n) q^n
\]

and is a modular function of weight zero on \( \text{SL}(2, \mathbb{Z}) \). Because of the holomorphy of \( \text{ch}_q(V) \) in \( H \) and our assumption that \( c \leq 24 \), one knows that (2.6) is uniquely determined up to an additive constant, and indeed is determined uniquely if \( c \leq 16 \). The upshot is this:

**Lemma 2.1.** **One of the following holds:**

\begin{enumerate}
  \item \( c = 8 \) and \( \text{ch}_q(V) = \Theta_{E_8}(q)/\eta(q)^8 = q^{-1/3}(1 + 248q + \cdots) \)
  \item \( c = 16 \) and \( \text{ch}_q(V) = (\Theta_{E_8}(q))^2/\eta(q)^{16} = q^{-2/3}(1 + 496q + \cdots) \)
  \item \( c = 24 \) and \( \text{ch}_q(V) = J(q) + \text{const} = q^{-1} + \text{const} + 196884q + \cdots \).
\end{enumerate}

Here, we have introduced the theta function \( \Theta_{E_8}(q) \) of the \( E_8 \) root lattice

\[
\Theta_{E_8}(q) = \sum_{\alpha \in E_8} q^{(\alpha, \alpha)/2}
\]

(where \( E_8 \) denotes the root lattice of type \( E_8 \) normalized so that the squared length of a root is 2) as well as the eta function

\[
\eta(q) = q^{1/24} \prod_{n \geq 1} \left( 1 - q^n \right);
\]
\[ J(q) = q^{-1} + 0 + 196884q + \cdots \]
is the absolute modular invariant normalized to have constant term zero (alias the graded character of the Moonshine Module \([FLM]\)). We also need the ‘unmodular’ Eisenstein series of weight two, namely

\[ E_2(q) = 1 - 24 \sum_{n \geq 1} \sigma_1(n) q^n \]

where \(\sigma_1(n)\) is the sum of the divisors of \(n\).

**Lemma 2.2.** For states \(u, v\) in \(V_1\) we have

\[
\text{Tr}_V o(u) o(v) q^{L(0) - c/24} = \langle u, v \rangle \left( \frac{24}{c} D_q(ch_q V) + E_2(q) ch_q V \right)
\]

where \(D_q = q \frac{d}{dq}\).

**Proof.** First recall the important identity, Proposition 4.3.5 in \([Z]\). If we pick a pair of states \(u, v \in V_1\) this result yields the following:

(2.7) \[
\text{Tr}_V o(u) o(v) q^{L(0) - c/24} = Z(u[-1]v, \tau) + 1/12 E_2(\tau) Z(u[1]v, \tau).
\]

The term \(Z(u[-1]v, \tau)\) is a modular form of weight 2, and we shall be able to write it down explicitly. Indeed, the l.h.s. of (2.7) has leading term \(\kappa(u, v) q^{1 - c/24}\) where \(\kappa(u, v)\) denotes the usual Killing form on \(V_1\), whereas the leading term of the second summand on the r.h.s. of (2.7) is equal to \(\frac{1}{12} \langle u, v \rangle q^{-c/24}\). From this discussion we conclude that the leading term of \(Z(u[-1]v, \tau)\) is equal to \(\frac{1}{12} \langle u, v \rangle q^{-c/24}\). We claim that up to scalars, the unique form of weight 2 on \(SL(2, \mathbb{Z})\) which is holomorphic in \(H\) and has a pole of order \(c/24\) at infinity is \(D_q(ch_q(V))\). To see this, suppose that \(f(\tau)\) and \(g(\tau)\) are two such forms. Then for some scalar \(\alpha\), \(f(\tau) - \alpha g(\tau)\) is a form of weight 2 on \(SL(2, \mathbb{Z})\) which is holomorphic both in \(H\) and at infinity. But it is well-known \([L]\) that such a form is identically zero, so that \(f(\tau) = \alpha g(\tau)\). So, up to scalars, there is at most one form \(f(\tau)\) with the stated properties. On the other hand, as \(ch_q V\) is a form of weight zero on \(SL(2, \mathbb{Z})\) which is holomorphic in \(H\) and has a pole of order \(c/24\) at infinity, then \(D_q(ch_q(V))\) is a weight 2 form \([L]\) with the same divisor. Thus \(D_q(ch_q(V))\) is weight 2 form on \(SL(2, \mathbb{Z})\) with the desired properties.

As a result, we see that the first term on the r.h.s. of (2.7) is equal to \(\frac{2}{c} \langle u, v \rangle D_q(ch_q(V))\). The lemma follows from this, (2.7), and (2.4). \(\square\)

Identifying coefficients of \(q^{1 - c/24}\) in the formula of Lemma 2.2 yields:

**Corollary 2.3.** \(\kappa(u, v) = 2 \langle u, v \rangle \left( \frac{\dim V_1}{c} - 1 \right)\).
3. Proof of the theorems.

We consider the three possibilities \( c = 8, 16, 24 \) in turn. In the first two cases the idea is to show that (II) *always* applies, while in the third case we study *when* it applies.

**Case 1.** \( c = 8 \). In this case \( \dim V_1 = 248 \) by Lemma 2.1 (a), so that \( \kappa(u,v) = 60\langle u,v \rangle \) by Corollary 2.3. Since \( \langle \cdot, \cdot \rangle \) is nondegenerate then so too is the Killing form. We conclude that in fact \( V_1 \) is semi-simple of dimension 248, and by (I) the Lie rank is no greater than 8. By the classification of semi-simple Lie algebras, we conclude that in fact \( V_1 \) is the Lie algebra of type \( E_8 \) and Lie rank 8. Now (II) and the fact that there is a unique positive-definite, even, unimodular lattice of rank 8, namely the \( E_8 \) root lattice, completes the proof of Theorem 1.

**Case 2.** \( c = 16 \). This is very similar to Case 1. Namely, we have \( \dim V_1 = 496 \) by Lemma 2.1 (b), whence \( \kappa(u,v) = 60\langle u,v \rangle \) once more. So again \( V_1 \) is semi-simple, and we proceed as in Case 1, now using the fact that there are just the two positive-definite, even, unimodular lattices. Theorem 2 follows.

**Case 3.** \( c = 24 \). Set \( \dim V_1 = d \). From Corollary 2.3 we see in this case that

\[
\kappa(u,v) = \langle u,v \rangle (d - 24)/12. \tag{3.1}
\]

If \( d = 24 \) then the Killing form is identically zero. Then \( V_1 \) is solvable by Cartan’s criterion and therefore abelian since \( V_1 \) is in any case reductive by (I). So if \( d = 24 \) then we have shown that \( V_1 \) has rank 24 and is therefore a lattice theory \( V_L \) for suitable \( L \) by (II). The fact that \( (V_L)_1 \) is abelian tells us that \( L \) has no roots i.e., no vectors of squared length two, and that \( L \) is therefore the Leech lattice (cf. [CS]). This deals with Case (b) of Theorem 3.

Let us assume from now on that \( d \neq 0, 24 \). Together with (3.1), this tells us that the Killing form on \( V_1 \) is again nondegenerate, so that \( V_1 \) is a semi-simple Lie algebra of Lie rank \( l \) no greater than 24 by (I). Moreover, if \( l = 24 \) then \( V \) is a lattice theory by (II) once more. This confirms Part (c) of Theorem 3. \( \square \)

Next we consider the levels of the affine Lie algebras spanned by the vertex operators \( Y(u,z) \), \( u \) in \( V_1 \). For states \( u, v \in V_1 \) and integers \( m, n \) we have

\[
[u_m, v_n] = (u_0v)_{m+n} + mu_1v_δ_{m,-n}, \tag{3.2}
\]

whereas the usual relations for a Kac-Moody Lie algebra of level \( k \) associated to a simple Lie algebra \( g \) take the form

\[
[a_m, b_n] = [a, b]_{m+n} + k(a, b)mδ_{m,-n} \tag{3.3}
\]

where \( (a, b) \) is the nondegenerate form on \( g \) normalized so that \( (\alpha, \alpha) = 2 \) for a long root \( \alpha \).
Let $V_1$ be a direct sum
\begin{equation}
V_1 = g_{1,k_1} \oplus g_{2,k_2} \oplus \cdots \oplus g_{n,k_n}
\end{equation}
of simple Lie algebras $g_i$, whose corresponding affine Lie algebra has level $k_i$. By comparing (3.2) and (3.3), using Corollary 2.3 we obtain for $u, v \in g_i$ that
\begin{equation}
\kappa_{g_i}(u,v) = (d - 24)k_i(u,v)/12
\end{equation}
where $\kappa_{g_i}$ denotes the restriction of the Killing form to $g_i$. Now $\kappa_{g_i}(h_\alpha, h_\alpha) = 4h_\alpha^\vee$ for a long root $\alpha$, where $h_\alpha^\vee$ is the dual Coxeter number of the root system associated to $g_i$. Therefore, (3.5) tells us that for each simple component $g_i$ of $V_1$, of level $k_i$ and dual Coxeter number $h_i^\vee$, the ratio
\begin{equation}
h_i^\vee / k_i = (d - 24)/24
\end{equation}
is independent of $g_i$.

**Proposition 3.1.** Assume that Case (a) of Theorem 3 holds. Then $B = V_2$ carries the structure of a (non-associative) simple, commutative algebra with respect to the product $a \cdot b = a_1b$.

**Proof.** We take for granted the well-known facts that $B$ is indeed a non-associative, commutative algebra with respect to the indicated product, and that the pairing $\langle \cdot, \cdot \rangle : B \times B \to \mathbb{C}$ defined by
\begin{equation}
a \cdot b = \langle a, b \rangle 1
\end{equation}
endows $B$ with a nondegenerate, invariant trace form. That is, $\langle \cdot, \cdot \rangle$ is symmetric and satisfies $\langle ab, c \rangle = \langle a, bc \rangle$. Moreover $B$ has an identity element $1/2\omega$. Set $d = \dim B = 196884$.

Next we state two more results that we will need. Each can be established using modular-invariance arguments along the same lines as before. Alternatively, we may use results of Section 4 of [M]:
\begin{equation}
\text{Tr}_{BO}(ab) = (d/3)\langle a, b \rangle.
\end{equation}
If $e^2 = e$ is in $B$ then
\begin{equation}
\text{Tr}_{BO}(e)^2 = 4620\langle e, e \rangle + 20336\langle e, e \rangle^2.
\end{equation}

Turning to the proof of the Proposition, we first show that $B$ is semisimple. Indeed, (3.8) guarantees that the form $\text{Tr}_{BO}(ab)$ is nondegenerate, and this is sufficient to establish that $B$ is semi-simple. To see this, recall a well-known result of Dieudonne (cf. [S, Theorem 2.6]) that an arbitrary algebra $B$ is semisimple if it has a nondegenerate trace form and contains no nonzero nilpotent ideals. We will show that indeed there are no nonzero nilpotent ideals in $B$. If not, we may choose a minimal nonzero nilpotent ideal $M$ in $B$. Note that $M^2 = 0$. Let $m$ be a nonzero element in $M$, and let $b$ in $B$ be arbitrary. Then $mb$ lies in $M$, so that $(mb)B \subset M$ and $(mb)M = 0$. This
shows that each element $mb$ is nilpotent as a multiplication operator on $B$, and hence $\text{Tr}_{B}mb = 0$ for all $b$. This contradicts the non-degeneracy of $\langle \cdot, \cdot \rangle$.

From the last paragraph we know that $B$ can be written as an (orthogonal) direct sum of simple ideals

$$B = B_1 + B_2 + \cdots + B_t.$$  

We must show that $t = 1$. Write $1/2\omega = e_1 + e_2 + \cdots + e_t$ where $e_i$ is the identity element of $B_i$. In particular, $e_i$ is a central idempotent of $B$. By (3.8) and (3.9) we obtain

$$\langle d/3, e, e \rangle = 4620\langle e, e \rangle + 20336\langle e, e \rangle^2$$

where $e$ is any of the idempotents $e_i$. Notice that (3.10) yields that

$$\langle e, e \rangle = 3.$$  

It is well-known that (3.11) implies that the components of the vertex operator $Y(2e, z)$ generate a Virasoro algebra of central charge 24, so that the total central charge must be 24$t$. Hence, $t = 1$ as required. □

### 4. Some Virasoro elements.

Throughout this section we assume that $V$ is a strongly rational, holomorphic vertex operator algebra of central charge 24 as before, and we assume in addition that $V_1$ is a (nonzero) semisimple Lie algebra of Lie rank $l$. Let $V_1$ have decomposition (3.4) into simple Lie algebras. Set $d_i = \dim g_{i,k_i}$. Let $(\cdot, \cdot)$ denote the normalized invariant bilinear form on $g_{i,k_i}$ with the property that $(\alpha, \alpha) = 2$ for a long root $\alpha$ in $g_{i,k_i}$. It is known (cf. [DL], [FZ], [K] and [L2]) that the element

$$\omega_i = \frac{1}{2(k_i + h_i^\vee)} \sum_{j=1}^{d_i} u_{j-1}^j u_{j-1}^{-1} 1$$

is a Virasoro element of central charge $c_i = \frac{k_i}{k_i + h_i^\vee} \dim g_{i,k_i}$, where $u^1, u^2, \ldots, u^{d_i}$ is an orthonormal basis of $g_i$ with respect to $(\cdot, \cdot)$.

There are three Virasoro elements in $V$ that are relevant to a further analysis of the situation. Namely, in addition to the original Virasoro element $\omega$ in $V$, we have

$$\omega_{aff} = \sum_{i=1}^{n} \omega_i$$

and

$$\omega_H = \frac{1}{2} \sum_{i=1}^{l} (h_{i-1}^1)^2 1$$
where \( h^1,\ldots,h^l \) is an orthonormal basis of a maximal abelian subalgebra \( H \) of \( V_1 \) with respect to the inner product \( \langle \cdot, \cdot \rangle \). Note that as a consequence of Equation (3.6), \( \omega_{aff} \) has central charge \( \sum c_i = 24 \). We omit further discussion of \( \omega_H \), but will prove:

**Proposition 4.1.** \( \omega_{aff} = \omega \).

**Proof.** Consequences of modular-invariance again underlie the proof of the Proposition, notably the absence of cusp-forms of small weight on \( SL(2,\mathbb{Z}) \). We introduce some notation for Virasoro operators: In addition to the usual operators \( L(n) \) associated to \( \omega \), we use \( L_{aff}(n) \) for the corresponding operators associated to \( \omega_{aff} \). We also set \( \omega' = \omega - \omega_{aff} \). We will soon see that \( \omega' \) is itself a Virasoro element, and define its component operators to be \( L'(n) \). We eventually want to prove of course that \( \omega' = 0 \). We proceed in a series of steps.

**Step 1.** \( \omega' \) is a highest weight vector of weight 2 for the Virasoro algebra generated by \( L(n) \).

The definition of \( \omega' \) shows that it lies in \( V_2 \), so it suffices to show that \( \omega' \) is annihilated by the operators \( L(1) \) and \( L(2) \). First calculate that if \( u \) is an element of \( V_1 \) then \([L(1),u_{-1}] = u_0\). Then from the definitions it follows easily that \( L(1) \) annihilates both \( \omega_{aff} \) and \( \omega_H \), whence it also annihilates \( \omega' \). To establish that \( L(2) \) annihilates \( \omega' \) we must show that

\[
L(2)\omega_{aff} = L(2)\omega_H = 12.
\]

Now for \( u \) in \( V_1 \) we have \([L(2),u_{-1}] = u_1\). Then

\[
L(2)u_{-1}^{i}u_{-1}^{j}1 = -\langle u^{j}, u^{i} \rangle = k_i,
\]

and (4.4) follows.

**Step 2.** \( \omega' \) is a Virasoro vector of central charge 0.

Use Step 1, in particular that \( L(1) \) annihilates \( \omega_{aff} \), to see that

\[
[L(m), L_{aff}(n)] = (m-n)L_{aff}(m+n) - 2(m^3 - m)\delta_{m,-n}.
\]

The result follows easily from this.

**Step 3.** \( Z_V(\omega',\tau) = 0 \).

Since \( \omega' \) is a highest weight vector of weight 2 for the Virasoro operators corresponding to \( \omega \) then from our earlier discussion of (2.5) we see that \( Z_V(\omega',\tau) \) is a modular form on \( SL(2,\mathbb{Z}) \) of weight 2 and is holomorphic in the upper half-plane. Moreover there is no pole in the \( q \)-expansion of \( Z_V(\omega',\tau) \), so \( Z_V(\omega',\tau) \) is in fact a holomorphic modular form of weight 2 on \( SL(2,\mathbb{Z}) \), hence must be zero.

**Step 4.** \( \text{Tr}_V L'(0)^2 q^{L(0)-1} = 0 \).
By Proposition 4.3.5 in [Z] we have
\[ \text{Tr}_{V'} L'(0)^2 q^{L(0)-1} = Z_V(L'[-2] \omega', \tau) - \sum_{k \geq 1} E_{2k}(\tau) Z_V(L'[2k-2] \omega', \tau) \]
where \( Y[\omega', z] = \sum_{m \in \mathbb{Z}} L'[m] z^{-m-2} \) and the functions \( E_{2k}(\tau) \) are Eisenstein series of weight \( 2k \), normalized as in [DLM3]. \( E_{2k}(\tau) \) is a holomorphic modular form on \( SL(2, \mathbb{Z}) \) if \( k > 1 \). Since \( L'[2k-2] \omega' = 0 \) if \( k \geq 2 \), \( L'[0] \omega' = 2 \omega' \), and \( L[0] L'[-2] \omega' = 4 L'[-2] \omega' \), we see that
\[ \text{Tr}_{V'} L'(0)^2 q^{L(0)-1} = Z_V(L'[-2] \omega', \tau) \]
is a modular form of weight 4 for \( SL(2, \mathbb{Z}) \). Since \( L'(0)V_1 = 0 \), it is in fact a cusp form, hence equal to zero.

**Step 5.** All the eigenvalues of \( L^{aff}(0) \) on \( V \) are real.

In order to see this recall the decomposition (3.4) and set \( \mathfrak{g} = V_1 \). Consider the affine Lie algebra
\[ \hat{\mathfrak{g}} = \mathfrak{g} \otimes \mathbb{C}[t, t^{-1}] \oplus \mathbb{C} \]
with bracket
\[ [x \otimes t^p, y \otimes t^q] = [x, y] \otimes t^{p+q} + p \delta_{p+q,0} (x, y) \]
for \( x, y \in \mathfrak{g} \) and \( p, q \in \mathbb{Z} \). Since each \( V_m \) is finite dimensional we see that \( V \) has a composition series as a module for \( \hat{\mathfrak{g}} \) such that each factor is an irreducible highest weight \( \hat{\mathfrak{g}} \)-module. So it is enough to show that \( L^{aff}(0) \) has only real eigenvalues on any irreducible highest weight \( \hat{\mathfrak{g}} \)-module. Note that such an irreducible highest weight \( \hat{\mathfrak{g}} \)-module is a tensor product of irreducible highest weight \( \hat{\mathfrak{g}}_i \)-modules \( L(k_i, \Lambda_i) \) of level \( k_i (i = 1, \ldots, n) \) for some dominant weight \( \Lambda_i \) in the weight lattice of \( \mathfrak{g}_i \) as \( V \) is a completely reducible \( \mathfrak{g}_i \)-module. Here \( L(k_i, \Lambda_i) \) is the unique irreducible quotient of the generalized Verma module \( U(\mathfrak{g}_i) \otimes_U L(\mathfrak{g}_i \otimes \mathbb{C}[t]) \otimes L(\Lambda_i) \) where \( L(\Lambda_i) \) is the highest weight module for \( \mathfrak{g}_i \) with highest weight \( \Lambda_i \) and \( x \otimes t^m \) acts as zero if \( m > 0 \) and \( x \otimes t^0 \) acts as \( x \). So it is enough to show that \( L_i(0) \) has only real eigenvalues on \( L(k_i, \Lambda_i) \) where \( Y(\omega_i, z) = \sum_{m \in \mathbb{Z}} L_i(m) z^{-m-2} \).

It is well-known that the eigenvalues of \( L_i(0) \) on \( L(k_i, \Lambda_i) \) are the numbers \( \frac{\langle \Lambda_i + 2 \rho_i, \Lambda_i \rangle}{2(k_i + h_i^\vee)} + m \) (cf. [DL] and [K]) for nonnegative integers \( m \) where \( \rho_i \) is the half-sum of the positive roots of \( \mathfrak{g}_i \). Since \( k_i \) is rational, it is clear that \( \frac{\langle \Lambda_i + 2 \rho_i, \Lambda_i \rangle}{2(k_i + h_i^\vee)} + m \) is real, as required.

**Step 6.** \( \omega' = 0 \).

By Step 5, all eigenvalues of \( L'(0) \) are real on \( V \). This implies that \( L'(0)^2 \) has only nonnegative eigenvalues. By Step 4 we conclude that all the eigenvalues of \( L'(0) \) are zero. Since \( L'(0) \omega' = 2 \omega' \), we must have \( \omega' = 0 \), as desired. \( \square \)
Added in proof. The authors have recently been able to establish the conjectured integrality of the levels. Details will appear elsewhere.

References


Received March 18, 2002 and revised April 1, 2003. The first author was supported by NSF grant DMS-9987656 and faculty research funds granted by the University of California at Santa Cruz. The second author was supported by NSF grant DMS-9700909 and faculty research funds granted by the University of California at Santa Cruz.

Department of Mathematics
University of California, Santa Cruz
Santa Cruz, CA 95064
E-mail address: dong@math.ucsc.edu

Department of Mathematics
University of California, Santa Cruz
Santa Cruz, CA 95064
E-mail address: gem@cats.ucsc.edu
PACIFIC JOURNAL OF MATHEMATICS
Vol. 213, No. 2, 2004

CORRESPONDENCES ON SHIMURA CURVES AND
MAZUR’S PRINCIPLE AT $p$

Frazer Jarvis

In this paper, we continue our earlier study of Mazur’s Principle for totally real fields, and extend the main result to one also applicable to primes dividing the characteristic. The method we use is a naive generalisation of that of Mazur and Ribet.

1. Introduction.

Let $F$ be a totally real field of degree $d$ over $Q$, and let $\ell$ be an odd prime. We recall the main result of [12], as improved by Fujiwara’s unpublished work [10]:

**Theorem 1.1** (Mazur’s Principle). Assume that

$$\rho : \text{Gal}(\overline{F}/F) \rightarrow \text{GL}_2(\mathbb{F}_\ell)$$

is a continuous irreducible semisimple representation which is attached to a Hilbert cuspidal eigenform $f \in S_{k,w}(U_0(p) \cap U_1(n))$, where $p \nmid n\ell$ and $k \geq 2t$. Suppose $[F(\mu_{\ell}) : F] \geq 4$. Then if $\overline{\rho}$ is irreducible, and unramified at $p$, and $N_{F/Q}(p) \not\equiv 1 \pmod{\ell}$, there exists a Hilbert cuspidal eigenform $f' \in S_{k,w}(U_1(n))$ to which $\overline{\rho}$ is attached.

Under an additional hypothesis, the same result is true when $[F(\mu_{\ell}) : F] = 2$. In [12], we proved this result with an additional hypothesis if $[F : Q]$ is even; this was removed in [10]. When $[F : Q]$ is odd, the results of this paper do not depend on unpublished work, but we need Fujiwara’s results when $[F : Q]$ is even.

Subsequently, Rajaei removed the restriction that $N_{F/Q}(p) \not\equiv 1 \pmod{\ell}$, thus generalising Ribet’s theorem, at least when $[F : Q]$ is odd. These results allow one to completely lower the level away from $\ell$, using other results of Fujiwara, as well as [13]. Further, Skinner and Wiles ([20]) have given a simple proof of level lowering if one allows the base field to be replaced with a suitable soluble extension.

The above theorem treats all cases with $\ell$ odd. If $\ell = 2$, then it never applies, as $p \nmid \ell$, which would imply that $N_{F/Q}(p)$ is odd, so certainly is congruent to 1 (mod 2). However, Buzzard ([3]) has a beautiful argument
which deals with mod 2 representations, and one can show ([14]) that it
generalises to totally real fields.

The behaviour when \(\ell = p\), the prime below \(p\), remains less clear, however. As
announced in [10], we may always assume that a modular mod \(p\) Galois
representation arises from a modular form of weight 2\(\ell\) (in Hida’s notation
[11]). For this reason, we will focus on weight 2\(\ell\) in what follows. This
allows us to work with Jacobians in a similar way to Ribet ([19]).

The main result of this paper (Theorem 6.2) is a partial result in the case
\(\ell = p\). In the same way as Ribet ([19]), we explain how to deduce a result
along the lines of “Shimura-Taniyama-Weil implies Fermat” for totally real
fields, assuming certain hypotheses on the field (which seem to be satisfied
only very rarely in practice).

It would be interesting to translate the proof of the main result of this
paper into the language of sheaf cohomology, and use techniques in \(p\)-adic co-
homology to deduce the same result directly for arbitrary weights. However,
the action of Galois on the cohomology of sheaves with \(p\)-power torsion on
curves with semistable reduction mod \(p\) still seems not to be completely un-
derstood, despite impressive recent work by Breuil, Fontaine, Hyodo, Kato
and Tsuji, amongst others. Indeed, our first attempts to solve this problem
followed this strategy; we may return to it at a later date.

The current paper may presumably also be viewed as one of the first steps
in the study of optimal weights for representations \(\overline{p}\).

It is a pleasure to thank the referee for many useful comments.

2. Shimura curves and integral models.

The proof of the above theorem, as well as the stronger version of Rajaei
([18]), is geometric in nature, involving a study of the cohomology of certain
Shimura curves. This approach was motivated by the study of Carayol ([5]).
We try to extend the method to give results for primes dividing \(p\).

If \(v\) denotes a prime of \(F\), then we let \(O_{F,v}\) denote the \(v\)-integral elements
of \(F\), and let \(O_{F,v}\) denote its completion, with residue field \(\kappa_v\).

Let \(B\) be a quaternion algebra over \(F\), ramified at all but one infinite
place \(\tau\), and not at any places above \(p\). We fix a ring isomorphism \(O_{B,v} \cong
M_2(O_{F,v})\) at all finite places \(v\) which split \(B\) and an isomorphism \(B_v \cong
M_2(F_v)\) at the unramified infinite place. The multiplicative group \(B^\times\) de-
nies, by restriction of scalars, an algebraic group over \(\mathbb{Q}\) denoted by \(G\).

Let \(A^\infty\) denote the finite adeles (over \(\mathbb{Q}\)). Associated to any open compact
subgroup \(K\) of \(G(A^\infty)\) is the Shimura curve

\[
M_K(\mathbb{C}) = G(\mathbb{Q}) \backslash (G(A^\infty)/K) \times X).
\]

Here, as with all our remarks on Shimura curves, the notation follows that of
[4], so that \(X\) denotes \(\mathbb{C} - \mathbb{R}\). By work of Shimura, there exists an \(F\)-scheme
MAZUR'S PRINCIPLE AT $p$

$M_K$ whose complex points are precisely $M_K(\mathbb{C})$ (where we regard $F$ as a subfield of $\mathbb{C}$ using the unramified infinite place $\tau$ of $B$).

In [5] and [12], it was necessary to study the cohomology of the special fibre of these curves, and it was thus necessary to prove the existence of models with certain properties for these curves defined over (the localisations of) the ring of integers $\mathcal{O}_F$ of $F$. Although it is likely that these exist in greater generality, this has only been demonstrated for certain subgroups $K$. We state the main results of [4] and [12].

Let $p$ denote a prime of $F$ above the rational prime $p$. Write $\kappa$ for the residue field $\kappa_p$. We suppose that $K$ factors as $K_p H$, where $H$ is an open compact subgroup of $\Gamma$, the restricted direct product over all finite places $v \neq p$ of $(B \otimes F_v)^\times$, as in [4], 0.4. Then we have the following results from [4] ((1) was already known to Morita):

**Theorem 2.1** (Carayol).

1) Suppose $K_p$ is the subgroup $K_p^0 = \text{GL}_2(\mathcal{O}_{F,p})$ (under the identification made above). Then, if $H$ is sufficiently small, there exists a model $\mathcal{M}_{0,H}$ of $M_K$ defined over $\mathcal{O}_{F,(p)}$. This model is proper and smooth over $\text{spec} \mathcal{O}_{F,(p)}$.

2) Suppose $K_p$ is the subgroup $K_p^n$ of matrices congruent to $I$ modulo $p^n$, again using the above identification. Then, if $H$ is sufficiently small (which depends on $n$), there exists a regular model $\mathcal{M}_{n,H}$ of $M_K$ with a map to $\mathcal{M}_{0,H}$. The morphism $\mathcal{M}_{n,H} \rightarrow \mathcal{M}_{0,H}$ is finite and flat.

We will also need to think about one other case, already considered in [12]:

**Theorem 2.2.**

1) Suppose $K_p$ is the subgroup

$$U_0(p) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(\mathcal{O}_{F,p}) \mid c \in p \right\},$$

again using the above identification. Then, if $H$ is sufficiently small (the same condition as 2.1(1)), there exists a regular model $\mathcal{M}_{U_0(p),H}$ of $M_K$ defined over $\mathcal{M}_{0,H}$. The morphism $\mathcal{M}_{U_0(p),H} \rightarrow \mathcal{M}_{0,H}$ is finite and flat.

2) The special fibre $\mathcal{M}_{U_0(p),H} \times \pi$ is isomorphic to a union of two copies of $\mathcal{M}_{0,H} \times \pi$ intersecting transversely above a finite set of points $\Sigma_H$.

The set of points $\Sigma_H$ are the supersingular points of $\mathcal{M}_{0,H} \times \pi$, and we use the same notation $\Sigma_H$ for the points which lie above them in $\mathcal{M}_{U_0(p),H} \times \pi$, the singular points of the special fibre. We now describe this set $\Sigma_H$. Let $\mathcal{B}$ denote the quaternion algebra got from $B$ by changing the invariants at $p$ and at $\tau$ (so it is now ramified at both these places, and is totally definite). Let $\mathcal{G}$ denote the algebraic group $\text{Res}_{F/Q} \mathcal{B}^\times$, and fix, for all places $v \neq p, \tau$,
an isomorphism between $B \otimes F_v$ and $\mathcal{B} \otimes F_v$. Then $\mathcal{C}(\mathbb{A}^\infty)$ may be identified with $\Gamma \times \mathcal{B}_p^\times$. By [4], 11.2(3), there is a bijection
\[
\Sigma_H \cong \mathcal{C}(\mathbb{Q}) \backslash \mathcal{C}(\mathbb{A}^\infty) / H \times \mathcal{O}_{\mathcal{B}_p}^\times
\cong \mathcal{C}(\mathbb{Q}) \backslash \Gamma \times F_p^\times / H \times \mathcal{O}_{F_p}^\times
\]
where the second isomorphism is induced by the reduced norm $\mathcal{B}_p^\times \rightarrow F_p^\times$.

Then $\Sigma_H$ is a finite set, and we denote its cardinality by $s_H$.

According to [4], 11.1.1, the action of $\text{GL}_2(F_p)$ on the inverse system of Shimura curves descends to an action on the supersingular points $\Sigma_H$, and the action factors through $\det : \text{GL}_2(F_p) \rightarrow F_p^\times$ (use the second description above of $\Sigma_H$ as a quotient). Further, if we normalise the reciprocity map of class field theory so that arithmetic Frobenius elements correspond to uniformisers (the opposite to [4]), then [4], 11.2(2), shows that an element $\sigma \in W(F_{ab}^p/F_p)$ acts on the set $\Sigma_H$ in exactly the same way as the element $[\sigma]$ of $F_p^\times$ corresponding to $\sigma$ by class field theory.

We recall from [12] that above each geometric point $x$ of $M_{0,H}$ was a divisible $\mathcal{O}_p$-module $E_{\infty|_x}$. The study of the integral model $M_{U_0(p),H}$ in characteristic $p$ arose from classifying isogenies of divisible $\mathcal{O}_p$-modules from $E_{\infty|_x}$ whose kernel was a group (scheme) isomorphic to $(\mathcal{O}_p/p)$. There were, in general, two families (which coincided at supersingular points), those arising from the Frobenius and those arising from the Verschiebung; this showed that $M_{U_0(p),H} \otimes \mathcal{O}$ consisted of two copies of $M_{0,H} \otimes \mathcal{O}$ which intersected above the supersingular points. Above the point $x \in M_{0,H} \otimes \mathcal{O}$ on one copy is the point $x$, together with the map $F : E_{\infty|_x} \rightarrow E_{\infty|_{\sigma x}}$, and on the other copy is the point $\sigma x$, together with the map $V : E_{\infty|_{\sigma x}} \rightarrow E_{\infty|_x}$.

3. Correspondences.

We are going to define correspondences on the $\mathbb{C}$-points of these Shimura curves. These maps will, in fact, be defined over $F$. These will then give rise to maps on the Jacobians, again defined over $F$; the Néron mapping property will then show that these maps extend to the Néron models of the Jacobians, and thus to their fibres. The only difference between this section and the discussion of correspondences in [19], is that in the setting of Shimura curves over more general totally real fields, the moduli interpretation of the correspondences used in [19] is not applicable, and instead we use an adelic approach.

We first recall that if $h \in G(\mathbb{A}^\infty)$, and $K_1$ and $K_2$ are open compact subgroups of $G(\mathbb{A}^\infty)$ such that $K_1 \subseteq hK_2h^{-1}$, then there is a covering map
\[
h : M_{K_1}(\mathbb{C}) \rightarrow M_{K_2}(\mathbb{C})
\]
\[
(g, x) \mapsto (gh, x)
\]
for $g \in G(\mathbb{A}^\infty)$, $x \in X$. By Shimura’s theory of canonical models, these maps come from maps on the underlying $F$-schemes. We will define the maps on the $\mathbb{C}$-points, but regard them as defined over $F$.

We first define the Hecke correspondence $T_p$. We use the same letter for a uniformiser of a prime ideal as for the ideal itself; thus $p$ will also denote a uniformiser for the prime ideal $p$. Let $\pi \in G(\mathbb{A}^\infty)$ denote the element which is the identity at every place apart from $p$, where it is $\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$.

We define the following two degeneracy maps between the curves $M_{U_0(p^r),H}$ (where we regard $U_0(p^0)$ as $\text{GL}_2(O_F,p)$):

$1 : M_{U_0(p^{r+1}),H} \rightarrow M_{U_0(p^r),H}$

$(g, x) \mapsto (g, x)$

for $g \in G(\mathbb{A}^\infty)$, $x \in X$, and

$\pi : M_{U_0(p^{r+1}),H} \rightarrow M_{U_0(p^r),H}$

$(g, x) \mapsto (g\pi, x)$. Then we let $T_p$ denote the map $\pi_*1^*$ from divisors on $M_{U_0(p^r),H}$ to themselves. This map induces an endomorphism of the Jacobian of the curve, which we denote by $J_{U_0(p^r),H}$. A lengthy, but entirely elementary, calculation shows that if $F = \mathbb{Q}$, $p = p$ and $B = \text{GL}_2$, then the map $T_p$ just defined is exactly the same map as that denoted $T_p$ in [19], pp. 443–4, the $p$th Hecke operator induced by Picard functoriality. (The map $1_*\pi^*$ is the Albanese version, denoted $\xi_p$ in [19].) We have thus defined the Hecke correspondence

$T_p : J_{U_0(p^r),H} \rightarrow J_{U_0(p^r),H}$. In the case $r = 0$, we write $\alpha = 1$ and $\beta = \pi$.

Next we define an Atkin-Lehner automorphism $w_p$ of $J_{U_0(p),H}$. For this, we let $w_p$ denote the element of $G(\mathbb{A}^\infty)$ which is the identity at every place apart from $p$, where it is $\begin{pmatrix} 0 & 1 \\ -p & 0 \end{pmatrix}$. As this element of $G(\mathbb{A}^\infty)$ normalises $U_0(p) \times H$, the map $(g, x) \mapsto (gw_p, x)$ gives a map from $M_{U_0(p),H}$ to itself, and thus an endomorphism of $J_{U_0(p),H}$. We denote this map also by $w_p$.

**Proposition 3.1.** Considered as maps from $J_{U_0(p),H}$ to itself, there is an equality

$\alpha^*\beta_* = T_p + w_p$.

**Proof.** This is a calculation that will be performed entirely on $\mathbb{C}$-points, but as all maps are defined over $F$, this will give the result.
Recall that
\[
M_{U_0(p^2), H} \rightarrow M_{U_0(p), H}
\]
\[
1 : (g, x) \mapsto (g, x)
\]
\[
\pi : (g, x) \mapsto (g\pi, x)
\]
and \(T_p = \pi_1^*\). Clearly
\[
T_p(g, x) = \sum_h (gh\pi, x),
\]
where \(U_0(p) = \coprod_h hU_0(p^2)\). (Abusively, we are using \(h\) to denote an element of \(U_0(p)\) as well as the corresponding element of \(G(A_{\infty})\) which is the identity at all other components.) It is easy to see that we may let \(h\) run through the set
\[
\left\{ \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \right\},
\]
where \(\alpha\) runs through a set of representatives in \(O_{F, p}\) of \(O_{F, p}/p\). It follows that
\[
T_p(g, x) = \sum_g \left( g \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} , x \right).
\]
Similarly,
\[
\alpha^*\beta_*(g, x) = \sum (g\pi h, x),
\]
where \(GL_2(O_{F, p}) = \coprod_h hU_0(p)\). We may let \(h\) run over the set
\[
\left\{ \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\},
\]
where \(\alpha\) runs through the set of representatives in \(O_{F, p}\) of \(O_{F, p}/p\) as above. It follows that
\[
\alpha^*\beta_*(g, x) = \sum \left( g \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} , x \right) + \left( g \begin{pmatrix} 0 & 1 \\ -p & 0 \end{pmatrix} , x \right).
\]
Comparing the two expressions, and recalling the definition of \(w_p\), the proposition follows. \(\square\)

If \(H\) is a subgroup which can be written as \(\prod_{q \neq p} H_q\), then we can define correspondences \(T_q = \begin{pmatrix} 1 & 0 \\ 0 & q \end{pmatrix} \) and \(S_q = \begin{pmatrix} q & 0 \\ 0 & q \end{pmatrix} \) in the same way, whenever \(H_q = GL_2(O_{F, q})\). The two maps in the first of these definitions are defined in the same way to those above; in the second, however, they are endomorphisms of the Jacobian of \(M_{U_p, H}\).
4. Hecke operators and Eisenstein ideals.

Suppose that \( S \) is a finite set of places of \( F \) containing all infinite places. We will introduce an action of a Hecke algebra \( T^S \), defined as the polynomial ring over \( \mathbb{Z} \) generated by \( T_q \) and \( S_q \) for primes \( q \notin S \). If \( p \in S \), the Hecke algebra \( T^{S,p} \) will be the \( \mathbb{Z} \)-algebra defined as above but with the additional generator \( T_p \).

Suppose \( S \) also contains all primes over which \( B \) is ramified. Then \( T^S \) acts naturally, through a quotient which we will denote by \( T^S_{k,w,B}(U) \), on the space \( S_{k,w,B}(U) \) of cusp forms of weight \((k,w)\) for any open compact subgroup \( U = \prod_q U_q \) of \( G(\mathbb{A}^\infty) \) such that \( U_q = \text{GL}_2(O_{F,q}) \) for all \( q \notin S \).

(All of our notation for Hilbert modular forms follows Hida [11] and our earlier papers [12] and [13].) We say that \( U \) is an \( S \)-subgroup when this holds.

We recall from [13] the definition of Eisenstein ideals in these Hecke algebras.

If \( m \) is a maximal ideal of \( T^S \) of residue characteristic \( p \), there is a homomorphism

\[
\overline{\theta}_m : T^S \longrightarrow T^S/m \hookrightarrow \overline{\mathbb{F}}_p,
\]

and an eigenform whose reduction has Hecke eigenvalues given by \( \overline{\theta}_m \); the semisimplification of the reduction of the associated Galois representation gives a representation

\[
\overline{\rho}_m : \text{Gal}(\overline{F}/F) \longrightarrow \text{GL}_2(\overline{\mathbb{F}}_p).
\]

In fact, the representation has a model valued in the field generated by the traces; as \( \text{tr} \overline{\rho}_m(\text{Frob}_q) = \overline{\theta}_m(T_q) \) for all but finitely many \( q \), the image of \( \overline{\rho}_m \) may be taken to be \( \text{GL}_2(T^S/m) \).

**Definition 4.1.** We say that \( m \) is Eisenstein if and only if \( \overline{\rho}_m \) is reducible. Equivalently, there should be some integral ideal \( \mathfrak{f} \) such that for all but finitely many primes \( q \) which are trivial in the narrow ray class group \( \text{Cl}(\mathfrak{f}) \), one has \( T_q - 2 \in m \) and \( S_q - 1 \in m \).

For more about Eisenstein ideals (and an explanation of the equivalence of the two definitions above), see [13], §3. Finally, we say that a \( T^S \)-module is Eisenstein if all maximal ideals in its support are Eisenstein.

Similar notions exist if \( T^S \) is replaced by \( T^{S,p} \).

5. Jacobians and Hecke operators.

In this section, we briefly outline the correspondence between Jacobians and the étale cohomology of the Shimura curve \( M_f \). We will then use the analysis of Carayol ([5]) to characterise the maximal ideals of the Hecke algebras which are in the support of the Jacobians.
Recall from [5], §2.3, that if $U$ is open compact in $G(\mathbb{A}^\infty)$, and $E \subset \mathbb{C}$ is a number field containing the Galois closure of $F$ and splitting $B$, then

$$H^1(M_U \otimes_F \mathbb{F}, E_\wp) \otimes_{E_\wp} E_\wp \cong \bigoplus_C (\pi_\infty)^U \otimes \rho_\pi(-1),$$

where $\wp$ is a prime of $E$, and $C$ denotes the set of cuspidal automorphic representations of $B$ of weight $k = 2t$. As $(\pi_\infty)^U$ has finite dimension, and is 0 for all but finitely many $\pi \in C$, this is therefore a finite direct sum of 2-dimensional $E_\wp$-vector spaces. The inductive system of these isomorphisms, as $U$ varies, is $G(\mathbb{A}^\infty)$-equivariant.

Now suppose that $U$ is an $S$-subgroup. Then the action of $G(\mathbb{A}^\infty)$ on the inductive system of cohomology groups leads, by a similar method to that outlined above for correspondences on Jacobians (see [11]), to an action of the Hecke algebra $T^S$ on $H^1(M_U \otimes_F \mathbb{F}, E_\wp)$. The above isomorphism shows that the maximal ideals of $T^S$ in its support correspond to cuspidal automorphic representations of weight $2t$ on $B$ with fixed vectors under $U$.

Choosing a Gal($\mathbb{F}/F$)-stable lattice for $\rho_\wp^\mathbb{F}$, and taking the semisimplification $\rho_\wp^\mathbb{F}$ of its reduction (in our application, the reduction will be irreducible, so taking the semisimplification will not be necessary), we see that $\rho_\wp^\mathbb{F}(-1)$ will be contained in

$$(H^1(M_U \otimes_F \mathbb{F}, \mathcal{O}_{E,\wp}) \otimes_{\mathcal{O}_{E,\wp}} \mathcal{O}_{E/\wp})^{ss};$$

by the argument of [12], §15, this space embeds into

$$(H^1(M_U \otimes_F \mathbb{F}, \mathcal{O}_{E/\wp}) \otimes_{\mathcal{O}_{E/\wp}} \mathcal{O}_{E/\wp})^{ss}.$$  

However, the cokernel is the $\ell$-torsion in the torsion free group $H^2(M_U \otimes_F \mathbb{F}, \mathcal{O}_{E,\wp})$, and so the embedding is actually an isomorphism. Moreover, $\mathcal{O}_{E/\wp}$ is flat over $\mathbb{Z}/p\mathbb{Z}$, and so this space is the same as

$$(H^1(M_U \otimes_F \mathbb{F}, \mathbb{Z}/p\mathbb{Z}) \otimes_{\mathbb{Z}/p\mathbb{Z}} \mathbb{F}_p)^{ss}.$$  

But it is well-known that for a smooth curve over an algebraically closed field of characteristic 0, the étale cohomology with coefficients in $\mu_p = \mathbb{Z}/p\mathbb{Z}(1)$ is canonically isomorphic to the $p$-torsion of the Jacobian of the curve.

It follows that $\rho_\wp^\mathbb{F}$ is a submodule of the $p$-torsion of the Jacobian of $M_U$, which we denote by $J_U$; in particular, the maximal ideals of $T^S$ in the support of $J_U[p]$ contain those corresponding to cuspidal automorphic representations on $B$ of weight $2t$ with fixed vectors under $U$, and conversely, any 2-dimensional irreducible submodule of (the semisimplification of) the $p$-torsion of $J_U$ is the reduction of the Galois representation associated to some cuspidal automorphic representation on $B$ of weight $2t$ with fixed vectors under $U$. 
In the case that \( U_p = U_0(p) \), we have analogously an action of \( T^{S,p}_k,w,B \), on the space \( S_k,w,B(U) \), the action of the operator \( T_p \) on the space of cusp forms being defined in the usual way ([11], p. 306).

If \( U \) is an \( S \)-subgroup, it follows that there is an action of the Hecke algebra \( T_S \) on \( J_U \) (using the Hecke correspondences \( T_q \) and \( S_q \) defined above) and that maximal ideals of the Hecke algebra in the support of \( J_U[p] \) correspond to automorphic representations of weight \( 2t \) on \( B \) which have fixed vectors under \( U \).

In the same way, if also \( U_p = U_0(p) \), the Hecke algebra \( T^{S,p}_k,w,B \) acts on \( J_U \), and similar results hold.


Definition 6.1. We say that a representation 
\[ \overline{\rho} : \text{Gal}(\overline{F}/F) \to \text{Aut}(V), \]
where \( V \) is a 2-dimensional \( \mathbb{F} \)-vector space (where \( \mathbb{F} \) is a finite field), is finite at \( p \) if the restriction of \( \overline{\rho} \) to a decomposition group at \( p \) is finite, i.e., if there exists a finite flat \( \mathbb{F} \)-vector space scheme \( V \) over \( \mathcal{O}_{F,p} \) such that the resulting representation of \( \text{Gal}(\overline{F}_p/F_p) \) coincides with \( \overline{\rho}|_{\text{Gal}(\overline{F}_p/F_p)} \).

We prove, in exactly the same way as [19], the following version of Mazur’s Principle:

Theorem 6.2 (Mazur’s Principle). Let \( F \) be a totally real number field, and let \( p \) be a prime of \( F \) dividing the rational prime \( p \). Suppose that 
\[ \overline{\rho} : \text{Gal}(\overline{F}/F) \to \text{GL}_2(\mathbb{F}_p) \]
is a continuous irreducible semisimple representation which is attached to a Hilbert cuspidal eigenform \( f \in S_{2t,1}(U_0(p) \cap U_1(n)) \), where \( p \nmid n \). Suppose also that:

1) If \( [F(\mu_p) : F] = 2 \), then \( \overline{\rho} \) is not induced from a character of \( \text{Gal}(\overline{F}/F(\sqrt{-3})) \).
2) \( e < p - 1 \), where \( e \) denotes the absolute ramification index of \( F_p \).

Then if \( \overline{\rho} \) is finite at \( p \), then there exists a Hilbert cuspidal eigenform \( f' \in S_{2t,1}(U_1(n)) \) to which \( \overline{\rho} \) is attached.

(Nota that we will throughout confuse \( U_0(p) \), defined as a subgroup of \( \text{GL}_2(\mathcal{O}_{F,p}) \), with the subgroup of \( G(\mathbb{A}^\infty) \) whose component at \( p \) is \( U_0(p) \), but which is the fixed maximal open compact subgroup at every place \( q \neq p \), and similarly for other subgroups defined only at a subset of all of the finite places.)

By the first assumption on \( F \), we may increase the level by replacing \( U_1(n) \) by \( H = U_1(n) \cap U_1^0(q_0) \) for some prime \( q_0 \nmid np \), in such a way that \( H \) is sufficiently small in the sense of Theorem 2.1 (and thus Theorem 2.2)
and that there are no congruences between forms of level $U_0(p) \cap U_1(n)$ and forms of level $U_0(p) \cap H$ which are new at $q_0$ in the sense that the eigenform does not arise from any level not a multiple of $q_0$. See [12], §12 for a proof of this fact.

Although the theorem above does not apply when $p = 2$, an analogous result on the existence of auxiliary primes is valid (as long as $\overline{p}$ is not induced from a character of $\text{Gal}(\mathbb{F}/F(\sqrt{-1}))$; for this, we can use a generalisation of an argument due to Kevin Buzzard which is proven in [14].

Let $T = T_{2t,t,\text{GL}_2}(U)$, where $U = U_0(p) \cap H$, and where $S$ denotes all primes dividing $nq_0$. Let $m$ denote the maximal ideal of $T$ corresponding to $\overline{p}$. Then there is a model of $\overline{p}$ defined over $T/m$.

Choose a prime $q_1 \nmid npq_0$ such that $\overline{p}(\text{Frob}_{q_1})$ is conjugate to $\overline{p}(\sigma)$, where $\sigma$ denotes complex conjugation. Then if $[F : Q]$ is even, the main result of [21] shows that $\overline{p}$ is also attached to a Hilbert cuspidal eigenform of level $U_0(pq_1) \cap H$ of weight $2t$ and which is new at $q_1$. (The same result is true for $[F : Q]$ odd by [18], Theorem 5, but we do not need this here; note, however, that this improves certain results in [13].)

If $[F : Q]$ is odd (resp. even), then let $B$ denote the quaternion algebra over $F$ ramified at $\{\tau_2, \ldots, \tau_d\}$ (resp. $\{q_1, \tau_2, \ldots, \tau_d\}$), and write $G$ for $\text{Res}_F/Q(B^\times)$. Let $T' = T_{2t,t,B}(U_0(p) \cap H)$ (resp. $T' = T_{2t,t,B}(U_0(pq_1) \cap U_0(pq_1) \cap H)$).

By [21], there is a maximal ideal $m'$ of $T'$ such that there is a natural isomorphism $\overline{p}_{m'} \equiv \overline{p}_m$.

We use the Jacquet-Langlands correspondence to see that then there is a cuspidal automorphic representation on $B$ of weight $2t$ corresponding to $f$ and with fixed vectors under $U_0(p) \cap H$. Let $m'$ denote the corresponding maximal ideal of $T'$, and let $F = T/m \cong T'/m'$. Let $V$ be the $2$-dimensional $F$-vector space corresponding to $\overline{p}$. By the results of the previous section, the corresponding maximal ideal of the Hecke algebra is in the support of the $p$-torsion of the Jacobian, $J_{U_0(p),H}[p]$.

Indeed, we see that $V$ is a $F(\text{Gal}(\mathbb{F}/F))$-submodule of $J_{U_0(p),H}[p]$. As $V$ is finite, it is the generic fibre of a finite flat $F$-vector space scheme $\mathcal{V}$ over $\mathcal{O}_{F,p}$.


After the preliminaries of the previous section, we now give the proof itself.

Lemma 7.1. The inclusion $\iota : V \hookrightarrow J$ prolongs to an embedding $\iota : \mathcal{V} \hookrightarrow \mathcal{J}$, where $\mathcal{J}$ is the Néron model of the Jacobian $J_{U_0(p),H}$ over $\mathcal{O}_{F,p}$.

The proof of this lemma is exactly the same as [19], Lemma 6.2.

Note that if $e = p - 1$, one should still be able to argue as in Edixhoven’s paper [9], Theorem 2.8; one needs to know that a Hilbert cuspidal eigenform of weight $2$ and level $pn$ ($p \nmid n$) which is not finite at $p$ must be ordinary.
at $p$. One should then be able to prove results like those of Wiles [22], Lemma 2.1.5, giving the structure of the local Galois representation at $p$. The remainder of the argument would follow as in [9]. (Thanks to Bas Edixhoven for explaining this argument.)

We recall from [1], [19] or SGA7, the structure of the special fibre $\overline{J}$ of the Néron model $J$ of $J_{U_0(p),H}$.

Combining the formalism of [19], §2, with Theorem 2.2 (note that $M_{U_0(p),H}$ has a regular integral model), we see that $\overline{J}^0$, the connected component of the special fibre of $J$, is an extension of the abelian variety $J_{0,H} \times J_{0,H}$ by a torus $T$ which one can describe explicitly ([19], Proposition 2.1). Let $X$ denote the character group of $\overline{T}$, and let $\Phi$ denote the group of components of $\overline{J}$ (as we have a regular integral model, an easy calculation shows that $\Phi \cong \mathbb{Z}/s_H\mathbb{Z}$). $\overline{T}$ lifts to a torus $T$ over $O_{F,p}$, which embeds in the formal completion of $J$ along $\overline{J}$.

**Lemma 7.2.** $T_p$ acts on $X$ in the same way as $\text{Frob}_p$.

**Proof.** Exactly as in [19], Proposition 3.7, we see that $T_p + w_p$ acts trivially on $T$, using Proposition 3.1 above. By the discussion at the end of §2, we know that $w_p$ acts on $\Sigma_H$ in the same way as $\text{Frob}_p$. It follows as in [19], Proposition 3.8, that $w_p$ acts on $X$ in the same way as $-\text{Frob}_p$. □

As in [19], Proposition 5.2(a), we see that $W = J_{U_0(p),H}[m]$ (semisimple by [2]), regarded as an $\mathbb{F}[\text{Gal}(\overline{F}/F)]$-module, decomposes as a product of copies of $V$, the Cartier dual of the group scheme given by $p$. For this, we note that $\overline{p}(\text{Frob}_q)$ satisfies $X^2 - S_q^{-1}T_qX + S_q^{-1}N_{\overline{F}/F}(q) = 0$ for all $q \notin S$; as $W \neq 0$ (because $m$ is in its support), the argument of [19] extends to this case.

Let us assume that $m$ is not in the support of $J_{0,H}[p]$. Let $\overline{V}$ denote the special fibre of $V$. As $V$ is contained in $J$, $\overline{V}$ is contained in $\overline{J}$. In our situation, the analogue of [19], Proposition 3.12, is trivial, because our integral model is regular (so, in the notation of [19], $\Lambda = \Lambda^*$, and so clearly $\eta_q(\Lambda^*) = \eta_q(\Lambda) \subset X$). It follows that $\overline{V}$ is contained in $\overline{J}^0$.

There is an extension

$$0 \longrightarrow \overline{T} \longrightarrow \overline{J}^0 \longrightarrow J_{0,H} \times J_{0,H} \longrightarrow 0.$$ 

By our assumption, $\overline{V}$ is contained in $\overline{T}$. The argument of [19], Lemma 6.3, extends to this case, and shows that $V$ is contained in $T$. It follows that

$$V \subset T[m](\mathbb{F}_p) = \text{Hom}(X/mX, \mu_p).$$

The Frobenius $\text{Frob}_p$ acts on $X/mX$ by $T_p$, using Lemma 7.2 above; this action is an automorphism, the negative of the automorphism $w_p$ defined above.
Let $I \subset \text{Gal}(F_p/F_p)$ denote the inertia group at $p$. Then $I$ acts on $V$ by $\chi$, the mod $p$ cyclotomic character.

As $V$ is 2-dimensional, the determinant of the action of $I$ on $V$ is given by $\chi^2$; however, we know that the determinant of $\rho$ is given by the cyclotomic character $\chi$, by the Čebotarev density theorem. Thus $\chi$ must be trivial on inertia. But $e < p - 1$, so $\mu_p \not\subset F_p$, and this implies that $\chi$ is not trivial on inertia. This contradiction shows that the assumption that $V$ is not modular of level $n$ is false. Note that if $[F : \mathbb{Q}]$ is even, we need to remove the auxiliary prime $q_1$ added at the end of §6 using Theorem 1.1. \hfill \Box

8. An application.

We end with an application of our main result, together with other known results on level lowering. The argument is exactly the same as Ribet’s argument in [19]: We attempt to deduce an analogue of Fermat’s Last Theorem over certain totally real number fields from the modularity of semistable elliptic curves. For a study of the (unfortunately rare!) occasions when the hypotheses of this theorem are satisfied in the case of real quadratic fields, see the work with Jayanta Manoharmayum ([15]) and Paul Meekin ([16] and [17]). For simplicity, we will assume that $F$ is a real quadratic field, although similar arguments are applicable for more general totally real fields.

We follow the idea of Frey, Serre and Ribet. Suppose that $\ell \geq 7$ is prime, and that $\alpha^\ell + \beta^\ell = \gamma^\ell$ is a point on the Fermat curve of degree $\ell$ and with $\alpha$, $\beta$ and $\gamma$ being nonzero elements of a totally real field $F$. Then we form the Frey curve $E$:

$$y^2 = x(x - \alpha^\ell)(x + \beta^\ell),$$

which is an elliptic curve over $F$. We ask whether this curve is modular, in the sense that its $L$-function agrees with the $L$-function of some weight 2 Hilbert cusp form over $F$. Let us make the following two hypotheses:

1) Suppose that $E$ is a semistable elliptic curve over $F$;
2) suppose that $\rho = \rho_{E,\ell}$ is absolutely irreducible.

As already remarked, we study these hypotheses over real quadratic fields in the papers [16] and [17]. Then we have the following application of the main theorem of the paper:

**Theorem 8.1.** Suppose (1) and (2) above. If the Frey curve $E$ is modular, then there is a Hilbert cuspidal eigenform

$$f \in S_{2\ell,\ell}(\Gamma^F_2)$$
such that $\mathfrak{p} \cong \mathfrak{p}_{f,\lambda}$ for some prime $\lambda$ of the number field generated by the Hecke eigenvalues of $f$, where

$$\Gamma_2^F = \bigcap_{p \mid 2} U_0(p) = U_0 \left( \prod_{p \mid 2} p \right).$$

The proof of this result is again exactly the same as that of Ribet [19]. One considers the mod $\ell$ representation $\mathfrak{p}$ associated to the Frey curve. The discriminant of the Frey curve is $16(\alpha \beta \gamma)^2 \ell$, and the conductor $n$ is the product of the primes dividing $\alpha \beta \gamma$ (as $E$ is semistable); as $E$ is modular, there is a Hilbert cuspidal eigenform $f'$ of weight $(2t, t)$ and level $n$ whose Galois representations coincide with those of $E$. From the theory of Tate curves, we see that if $p \nmid 2 \ell$ is a prime ideal of $\mathcal{O}_F$ dividing $n$, then $\mathfrak{p}$ is unramified at $p$. Further, if $p \mid \ell$, then one argues as in Proposition 8.2 of [9] that $\mathfrak{p}$ is finite at $p$.

As $[F : \mathbb{Q}]$ is even, we begin by adding an auxiliary prime to the level as in §8 of [13] (which we may do thanks to [21]). (This would not be necessary when $[F : \mathbb{Q}]$ is odd.)

Now we lower the level. As $[F : \mathbb{Q}] = 2$, then we see that if $\ell > 3$, we will certainly have $e < \ell - 1$. Indeed, $\ell \geq 7$, so that also $[F(\mu_\ell) : F] > 2$. Now we can use the result of this paper to remove all primes dividing $\ell$ from $n$. Next, we may use the results of [12] and [18] to remove the remaining primes not dividing 2 from the level. Finally, we can remove the auxiliary prime which we added by using Fujiwara’s version (Theorem 1.1) of Mazur’s Principle in the even degree case. The conclusion is that there is some (adelic) Hilbert cuspidal eigenform $f$ of weight $(2t, t)$ and of (semistable) level dividing 2, i.e., on the group given in the statement of the theorem.

As a final remark, we should point out that, as in the case of $\mathbb{Q}$, there is presumably a variant of the main result of this paper in terms of “lowering the weight”.

References


Received July 10, 2002 and revised May 12, 2003.

DEPARTMENT OF PURE MATHEMATICS
UNIVERSITY OF SHEFFIELD
SHEFFIELD S3 7RH
U.K.
E-mail address: a.f.jarvis@shef.ac.uk
HARNACK INEQUALITY FOR NONDIVERGENT ELLIPTIC OPERATORS ON RIEMANNIAN MANIFOLDS

Seick Kim

We consider second-order linear elliptic operators of nondivergence type which are intrinsically defined on Riemannian manifolds. Cabré proved a global Krylov-Safonov Harnack inequality under the assumption that the sectional curvature is nonnegative. We improve Cabré’s result and, as a consequence, we give another proof to the Harnack inequality of Yau for positive harmonic functions on Riemannian manifolds with nonnegative Ricci curvature using the nondivergence structure of the Laplace operator.

1. Introduction and main results.

In this article we study Harnack inequalities for solutions of second-order elliptic equations on Riemannian manifolds. The Harnack inequality is well understood for divergence form operators on manifolds satisfying certain properties, namely the volume doubling property and the weak Poincaré inequality; see [6, 10] and [11]. For example, the above two conditions hold for manifolds with nonnegative Ricci curvature. In fact, Grigor’yan [6] and Saloff-Coste [11] proved that the volume doubling property and the weak Poincaré inequality imply the Harnack inequality for solutions to divergence type parabolic equations. Furthermore, Saloff-Coste [11] showed that the above two conditions are equivalent to the Harnack inequality for the heat operator.

Recently, Cabré [1] considered nondivergence type elliptic operators and proved the Harnack inequality on manifolds with nonnegative sectional curvature; see below for the definition of nondivergent elliptic operators. His result is an extension of the Euclidean Krylov-Safonov Harnack inequality ([7] and [9]). Compared to the Harnack inequality for divergent operators, the hypothesis in the Harnack inequality of Cabré is not completely optimal and we improve it here. For example, the Laplace operator has both divergent and nondivergent structure and it is well-known that positive harmonic functions on Riemannian manifolds with nonnegative Ricci curvature satisfy the Harnack inequality, which was earlier proved by Yau [15]. Since the assumption in the Harnack inequality of Cabré requires the underlying...
manifold to have nonnegative sectional curvature, it does not directly imply
the Harnack inequality of Yau.

The motivation for our work was to give another proof of Yau’s Harnack
inequality using the nondonvrgent structure of the Laplace operator; and by
doing so, we expected to generalize Cabré’s result. We found a sufficient
condition in terms of the distance function on the underlying Riemannian
manifold, which implies a global Krylov-Safonov Harnack inequality for sol-
solutions of uniformly elliptic equations of nondonvrgence type. We show that
our result implies the Harnack inequality of Yau, and we will provide exam-
ple of such Riemannian manifolds, using the curvature assumption. Our
examples strictly include the manifolds with nonnegative sectional curva-
ture, which means that our result generalizes that of Cabré [1].

Let us begin with the definition of nondonvrgent elliptic operators on Rie-
mannian manifolds. Let \((M, g)\) be a smooth, complete, noncompact Rie-
mannian manifold of dimension \(n\). For any \(x \in M\), we denote by \(T_x M\) the
tangent space of \(M\) at \(x\). For any \(x \in M\), let \(A_x\) be a positive definite
symmetric endomorphism of \(T_x M\). We assume that
\[
\lambda |X|^2 \leq \langle A_x X, X \rangle \leq \Lambda |X|^2 \quad \forall x \in M, \quad \forall X \in T_x M,
\]
for some positive constants \(\lambda\) and \(\Lambda\). Here \(\langle X, Y \rangle = g(X, Y)\) and \(|X|^2 = \langle X, X \rangle\). We consider a second-order, linear, uniformly elliptic operator \(L\) defined by
\[
Lu = \text{tr}(A_x \circ D^2 u) = \text{tr}\{X \mapsto A_x \nabla_X \nabla u\},
\]
where \(\circ\) denotes composition of endomorphisms, \(\text{tr}\) is the trace, and \(D^2 u\)
denote the Hessian of the function \(u\). We recall that the Hessian of \(u\) at
\(x \in M\) is the endomorphism of \(T_x M\) defined by
\[
D^2 u \cdot X = \nabla_X \nabla u,
\]
where \(\nabla u(x) \in T_x M\) is the gradient of \(u\) at \(x\). It may seem that our definition
of nondonvrgent elliptic operator differs from that of Cabré [1], but since
\[
Lu = \text{tr}(A_x \circ D^2 u) = \text{tr}(D^2 u \circ A_x) = \text{tr}\{X \mapsto \nabla_{A_x X} \nabla u\},
\]
our definition of \(Lu\) coincides with the one that originally appeared in Cabré
[1]. We point out that Stroock also used the convention (2) of nondonvrgent
elliptic operators in his recent article [14].

In contrast, divergent operators in manifolds are of the form
\[
\mathcal{L}u = \text{div}(A_x \nabla u),
\]
where \(\text{div}\) denotes the divergence. The Laplace operator has both divergent
and nondonvrgent structure in the sense that \(\Delta u\) can be viewed as either
\(\Delta u = \text{div}(\nabla u)\) or \(\Delta u = \text{tr}(D^2 u)\).

Let \(p\) be a fixed point in \(M\) and let \(d_p(x)\) be the distance function defined
by \(d_p(x) = d(p, x)\), where \(d(p, x)\) is the geodesic distance between \(p\) and \(x\).
Recall that \( d_p \) is smooth on \( M \setminus (\text{Cut}(p) \cup \{p\}) \), where \( \text{Cut}(p) \) is the cut locus of \( p \); see e.g., \([12]\) and \([13]\). We assume that \( M \) satisfies the following conditions:

\[
\Delta d_p(x) \leq \frac{n-1}{d_p(x)} \quad \text{for} \quad x \notin \text{Cut}(p) \cup \{p\} \quad \forall p \in M \quad \text{and} \quad \forall p \in M,
\]

where \( a_L \) is a constant which may depend on the given operator \( L \).

By the Laplace comparison theorem (see e.g., \([12]\) and \([13]\)), any manifold with nonnegative Ricci curvature satisfies Condition (3). We shall give examples of manifolds satisfying Condition (4) shortly.

Let us state our main theorem. We would like to mention that the constant \( C \) which appears in the following main result does not depend on the smoothness of the coefficients \( A_x \). Indeed, it is one of the most important features in Moser and Krylov-Safonov type Harnack inequalities, including that of Cabrér [1], which remain valid for operators with bounded measurable coefficients.

**Theorem 1.1.** Assume that \( M \) satisfies Conditions (3) and (4). Let \( u \) be a smooth function in a ball \( B_{2R} \) satisfying \( u \geq 0 \) in \( B_{2R} \). Then

\[
\sup_{B_R} u \leq C \left\{ \inf_{B_R} u + \frac{R^2}{|B_{2R}|^{1/n}} \|Lu\|_{L^n(B_{2R})} \right\},
\]

where \( C \) is a constant depending only on \( \lambda, \Lambda, n \) and \( a_L \).

Here, \( B_{2R} \) is any geodesic ball of radius \( 2R \), \( B_R \) is the geodesic ball of radius \( R \) concentric with \( B_{2R} \), and \( |B_{2R}| \) is the volume of \( B_{2R} \). As an immediate consequence, we obtain the following Liouville property for solutions of \( Lu = 0 \) in \( M \):

**Corollary 1.2.** Let \( u \) be a smooth solution of \( Lu = 0 \) in \( M \) that is bounded from below. Then \( u \) is constant.

We now give examples of manifolds satisfying (3) and (4). This can be done in a most transparent way by stating it in terms of Pucci’s extremal operator. For a symmetric endomorphism \( S \) on \( T_x M \), we define

\[
\mathcal{M}^{-}[S, \lambda, \Lambda] = \mathcal{M}^{-}[S] = \lambda \sum_{\kappa_i > 0} \kappa_i + \Lambda \sum_{\kappa_i < 0} \kappa_i,
\]

where \( \kappa_i = \kappa_i(S) \) are the eigenvalues of \( S \). Suppose \( M \) satisfies the following curvature assumption: For any unit vector \( e \) in \( T_x M \) and for all \( x \in M \),

\[
\mathcal{M}^{-}[R(e)] \geq 0,
\]

where \( R(e) \) is the Ricci transformation on \( T_x M \); see Sec. 2 for its definition. Note that the inequality \( \mathcal{M}^{-}[R(e)] \leq \text{tr}(A_x \circ R(e)) \) holds provided that
$A_x$ satisfies the ellipticity condition (1). Thus, by taking $A_x = \lambda \text{Id}$, we find that Condition (5) implies that $M$ has nonnegative Ricci curvature and hence, by the Laplace comparison theorem (see [12] and [13]), it implies (3). We shall see in the next section that (5) also implies Condition (4) with $a_L = (n - 1)\Lambda$; see Lemma 2.1 below. Then, the next corollary will follow immediately.

**Corollary 1.3.** Assume that $M$ satisfies the curvature condition (5). Let $u$ be a smooth function in a ball $B_{2R}$ satisfying $u \geq 0$ in $B_{2R}$. Then

$$\sup_{B_R} u \leq C \left\{ \frac{R^2}{|B_{2R}|^{1/n}} \|Lu\|_{L^n(B_{2R})} + \inf_{B_{2R}} u + R^2 \right\},$$

where $C$ is a universal constant; i.e., a constant depending on the ellipticity constants $\lambda, \Lambda$ and the dimension $n$ only.

Note that in the case when $L = \Delta$, Condition (5) is just to say that the Ricci curvature is nonnegative. Hence, Theorem 1.1 and Corollary 1.2, respectively, imply the Harnack inequality and the Liouville property by Yau [15]. Moreover, since Condition (5) holds trivially for manifolds with nonnegative sectional curvature, our result also generalizes the Harnack inequality of Cabr´ e [1].

2. Preliminaries.

Recall, for $x \in M$, the exponential map $\exp_x : T_x M \to M$. For $X \in T_x M$ with $|X| = 1$, $\gamma(t) = \exp_x(tX)$ is the unique unit speed geodesic that starts from $x$ and goes in the direction of $X$. Let

$$t_0 = \sup \{ t > 0 : \gamma \text{ is the unique minimal geodesic joining } x \text{ and } \gamma(t) \}.$$

If $t_0 < \infty$, then $\gamma(t_0)$ is called a cut point of $x$. We denote

$$\text{Cut}(x) = \text{set of all cut points of } x.$$

If we denote $S_x = \{ X \in T_x M : |X| = 1 \}$, it is clear that for any $X \in S_x$, there can be at most one cut point on the geodesic $\exp_x(tX)$, $t > 0$. If $\exp_x(t_0X) = y$ is a cut point of $x$ then we set $\mu(X) = d(x, y)$, the geodesic distance between $x$ and $y$. If there is no cut point, we set $\mu(X) = \infty$. Define

$$E_x := \{ tX : 0 \leq t < \mu(X), X \in S_x \}.$$

Then, it can be shown that $\exp_x : E_x \to \exp_x(E_x)$ is a diffeomorphism and, also that $\text{Cut}(x)$ has $n$-dimensional measure zero; see [12] and [13].

Let the Riemannian curvature tensor be defined by

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z,$$

where $\nabla$ is the Levi-Civita connection. For a unit vector $e \in T_x M$, $R(e)$ will denote the Ricci transformation of $T_x M$ into itself given by $R(e)X :=$
$R(X,\varepsilon)e$. The Morse index form $I(\cdot, \cdot)$ is defined by

$$I(X, Y) = I_0(X, Y) = \int_0^\ell \left\{ \langle \nabla_{\gamma'}X, \nabla_{\gamma'}Y \rangle - \langle R(\gamma', X)Y, \gamma' \rangle \right\} \, dt,$$

where $\gamma : [0, \ell] \to M$ is a geodesic parametrized by arc length and $X, Y$ are piecewise smooth vector fields along $\gamma$.

Now, we show that Condition (5) implies (4) with $a_L = (n - 1)\Lambda$.

**Lemma 2.1.** Let $M$ be an $n$-dimensional complete Riemannian manifold satisfying the curvature condition (5). Let $d_p$ be the distance function from a fixed point $p \in M$. If $x \notin \text{Cut}(p) \cup \{p\}$, then

$$Ld_p(x) \leq \frac{a_L}{d_p(x)}, \quad \text{where} \quad a_L = (n - 1)\Lambda.$$

**Proof.** Let $\gamma : [0, \rho] \to M$ be the minimal geodesic parametrized by arc length joining $\gamma(0) = p$ and $\gamma(\rho) = x$. Choose an orthonormal basis $\{e_i\}_{i=1}^n$ on $T_xM$ such that $e_1 = \gamma'(0)$ and $\{e_i\}_{i=1}^n$ are eigenvectors of $D^2d_p$ on $T_xM$. We extend $\{e_i\}_{i=1}^n$ to $\{e_i(t)\}_{i=1}^n$ on $t \in [0, \rho]$ by parallel translation. Let $X_i$, $i = 2, \ldots, n$, be the Jacobi fields along $\gamma$ such that $X_i(0) = 0$, $X_i(\rho) = e_i$ and $[X_i, \gamma'] = 0$. Then $\langle D^2d_p(e_1), e_i \rangle = \langle \nabla_{\gamma'}X_i, X_i \rangle(\rho) = I(X_i, X_i)$. Let $Y_i = f(t)e_i(t)$, where $f(t) = \frac{2}{\rho}$.

Since a Jacobi field minimizes the index form among all vector fields along the same geodesic with the same boundary data, we have $I(X_i, X_i) \leq I(Y_i, Y_i)$. Denote $a_{ij} := \langle A_xe_i, e_j \rangle$. Then

$$Ld_p(x) = \sum_{i=2}^n a_{ii} \langle D^2d_p(e_i), e_i \rangle = \sum_{i=2}^n a_{ii}I(X_i, X_i) \leq \sum_{i=2}^n a_{ii}I(Y_i, Y_i).$$

On the other hand,

$$\sum_{i=2}^n a_{ii}I(Y_i, Y_i) = \sum_{i=2}^n a_{ii} \int_0^\rho |f'|^2 - \int_0^\rho f^2 \sum_{i=2}^n a_{ii} \langle R(\gamma', e_i)e_i, \gamma' \rangle$$

$$\leq \sum_{i=2}^n a_{ii} \int_0^\rho |f'|^2 - \int_0^\rho f^2 \mathcal{M}^-[R(\gamma')]$$

$$\leq \frac{(n - 1)\Lambda}{\rho}.$$

This completes the proof. \qed

Observe that in the case when $\lambda = \Lambda = 1$ (i.e., $L = \Delta$), Condition (5) reduces to the nonnegative Ricci curvature assumption on $M$. Also, note that if the sectional curvature of $M$ is nonnegative, then (5) holds trivially.

Let us recall that, if $\phi$ is a smooth map from $M$ to another Riemannian manifold $N$, the Jacobian of $\phi$ is the absolute value of the determinant of the differential of $\phi$, that is, $\text{Jac } \phi(x) := |\det d\phi(x)|$. This determinant is
computed when expressing $d\phi(x)$ in an orthonormal basis of $T_xM$ and an orthonormal basis of $T_{\phi(x)}N$, and hence it is defined up to a sign. Its absolute value, $\text{Jac} \phi(x)$, is therefore well-defined. The following is the area formula on $M$, which follows easily from the area formula in Euclidean space using a partition of unity; see [1]: For any smooth map $\phi$ from $M$ to $M$ and any measurable subset $E$ of $M$, we have

$$\int_E \text{Jac} \phi(x) \, dV(x) = \int_M \mathcal{H}^0[E \cap \phi^{-1}(y)] \, dV(y),$$

where $\mathcal{H}^0$ is the counting measure.

The following lemma is quoted from [1, Lemma 3.2].

**Lemma 2.2** (Cabré). Let $v$ be a smooth function in an open set $\Omega$ of $M$. Consider the map $\phi$ from $\Omega$ to $M$ defined by

$$\phi(p) = \exp_p \nabla v(p).$$

Let $x \in \Omega$ and suppose that $\nabla v(x) \in E_x$. Set $y = \phi(x)$. Then

$$\text{Jac} \phi(x) = \text{Jac} \exp_x(\nabla v(x)) \cdot |\det D^2(v + d^2 y/2)(x)|,$$

where $\text{Jac} \exp_x(\nabla v(x))$ denotes the Jacobian of $\exp_x$, a map from $T_xM$ to $M$ at the point $\nabla v(x) \in T_xM$.

In the normal polar coordinates $(r, \theta)$, the area element $J(r, \theta) \, d\theta$ of the geodesic sphere $\partial B_r(x)$ of radius $r$ centered at $x$ is given by $r^{n-1} A(r, \theta) \, d\theta$, where $A(r, \theta)$ is the Jacobian of the map $\exp_x$ at $r \theta \in T_xM$. Assume that $M$ satisfies Condition (3). For $y = (r, \theta)$ not in the cut locus of $x$, we have the inequality $\frac{\partial}{\partial r} \ln J(r, \theta) = \Delta r \leq \frac{n-1}{r} \geq \frac{\partial}{\partial r} \ln A(r, \theta)$; see [8]. By integrating this, we find $A(r, \theta) = r^{1-n} J(r, \theta)$ is a nonincreasing function of $r$ and, in particular, $A(r, \theta) \leq 1$ since $\lim_{r \to 0} A(r, \theta) = 1$. Hence, Lemma 2.2 implies

$$\text{Jac} \phi(x) \leq |\det D^2(v + d^2 y/2)(x)|.$$  

Also, from the above observation, it follows that $r^{-n} |B_r|$ is nonincreasing in $r$; see e.g., [8] and [13]. This is generally known as Bishop’s volume comparison theorem. We state it as a lemma.

**Lemma 2.3** (Bishop). Let $M$ be an $n$-dimensional Riemannian manifold satisfying (3). For any $x \in M$, $|B_R(x)|/R^n$ is nonincreasing with respect to $R$. In particular,

$$|B_R(x)|/|B_r(x)| \leq R^n/r^n \quad \text{if} \quad 0 < r < R,$$

and $M$ satisfies the volume doubling property; i.e., $|B_{2R}(x)| \leq 2^n |B_R(x)|$. 


Throughout the entire section, we shall assume that $M$ satisfies (3) and (4). We will closely follow the outline of Cabré’s proof of the Harnack inequality in [1], which in turn carried over the basic scheme in the book by Caffarelli and Cabré [3]; see also [2]. Our goal is to establish Lemma 3.1 below. It corresponds to [1, Lemma 5.1], in the proof of which Cabré used the assumption that the underlying manifold $M$ has nonnegative sectional curvature. If one investigates his arguments carefully, it will turn out that, besides [1, Lemma 5.1], the only geometrical property of $M$ he used in proving the subsequent lemmas in Sec. 6 and Sec. 7 of [1] is the volume growth condition (8). Therefore, once we prove Lemma 3.1, the proof of Theorem 1.1 follows from the one in [1] verbatim since, by Lemma 2.3 of the previous section, Property (8) holds for any manifold satisfying Condition (3).

**Lemma 3.1.** Let $u$ be a smooth function in a ball $B_{7R}$ satisfying

$$Lu \leq f \text{ in } B_{7R}, \quad u \geq 0 \text{ in } B_{7R}, \quad \inf_{B_{2R}} u \leq 1,$$

and

$$\frac{R^2}{|B_{7R}|^{1/n}} \| f \|_{L^n(B_{7R})} \leq \varepsilon_\delta.$$

Then, for any $0 < \delta < 1$,

$$\frac{|\{u \leq M_\delta \} \cap B_{5R}|}{|B_{7R}|} \geq \mu_\delta,$$

(9)

where $\varepsilon_\delta$, $0 < \mu_\delta < 1$ and $M_\delta > 1$ are positive constants depending on $\delta$.

We need a series of lemmas to prove Lemma 3.1 above. The next lemma is a modification of [1, Lemma 4.1], where Cabré assumed the nonnegativity of the sectional curvature to get the same conclusion. We point out that the Krylov-Safonov theory in Euclidean space is based on the Alexandrov-Bakelman-Pucci (ABP) estimate, which in turn is based on the existence of linear functions with constant gradient on Euclidean space. Such functions do not exist in most Riemannian manifolds and were replaced in [1] by paraboloids.

**Lemma 3.2.** Let $u$ be a smooth function in a ball $B_{7R} := B_{7R}(p)$ satisfying

$$u \geq 0 \text{ in } B_{7R} \setminus B_{5R} \text{ and } \inf_{B_{2R}} u \leq 1.$$  

Then

$$|B_R| \leq \frac{1}{(n\lambda)^n} \int_{\{u \leq 0\} \cap B_{5R}} \left\{ (R^2Lu + a_L + \Lambda)^+ \right\}^n dV.$$

Here, we denote by $f^+$ the positive part of a function, i.e., $f^+ = \max(f, 0)$. 

Proof. For any $y \in B_R$ we consider the continuous function
\[ w_y := R^2 u + \frac{1}{2} d^2_y. \]
We have that $\inf_{B_{2R}} w_y \leq R^2 + (3R)^2/2 = 11R^2/2$. In $B_{7R} \setminus B_{5R}$ we have, since $u \geq 0$ here, $w_y \geq (4R)^2/2 > 11R^2/2$. We conclude that the minimum of $w_y$ in $\overline{B_{5R}}$ is achieved at some point of $B_{5R}$, which is also a minimum of $w_y$ in $B_{7R}$. That is,
\[ \inf_{B_{7R}} w_y = \inf_{B_{5R}} w_y = w_y(x), \]
for some $x \in B_{5R}$. It is not hard to see that $y = \exp_x \nabla (R^2 u)(x)$; see [1, pp. 637–638]. We are therefore led to consider the smooth map from $B_{7R}$ to $M$
\[ \phi(z) = \exp_z \nabla (R^2 u)(z), \]
and the measurable set
\[ E := \left\{ x \in B_{5R} : \exists y \in B_R \text{ such that } w_y(x) = \inf_{B_{7R}} w_y \right\}. \]
We have proved that for any $y \in B_R$ there is at least one $x \in E$ such that $\phi(x) = y$. The area formula (6) gives
\[ |B_R| \leq \int_M \mathcal{H}^0[E \cap \phi^{-1}(y)] dV(y) = \int_E (\text{Jac } \phi) dV. \tag{11} \]
We claim that $\text{Jac } \phi(x) \leq (n\lambda)^{-n}\{R^2 Lu(x) + a_L + \Lambda\}^n$ for any $x \in E$.
Let $x \in E$ and take $y \in B_R$ such that $w_y(x) = \inf_{B_{7R}} w_y$. If $x$ is not a cut point of $y$, then by (7) we find
\[ \text{Jac } \phi(x) \leq \left| \det D^2(R^2 u + d^2_y/2)(x) \right| = \left| \det D^2 w_y(x) \right|. \]
Since $w_y$ achieves its minimum at $x$, $D^2 w_y(x) \geq 0$. Here we used $\leq$ to denote the usual order between symmetric endomorphisms. Therefore, by using the well-known inequality
\[ \det A \cdot \det B \leq \{n^{-1} \text{tr}(A \circ B)\}^n, \quad A, B \text{ symmetric } \geq 0, \]
we conclude
\[ \text{Jac } \phi(x) \leq \det D^2 w_y(x) \leq \frac{1}{\lambda^n} \det A_x \cdot \det D^2 w_y(x) \]
\[ \leq \frac{1}{(n\lambda)^n} \{\text{tr} \left( A_x \circ D^2 w_y(x) \right)\}^n \]
\[ = \frac{1}{(n\lambda)^n} \{L w_y(x)\}^n \]
\[ = \frac{1}{(n\lambda)^n} \{R^2 Lu(x) + L(d^2_y/2)(x)\}^n \]
\[ \leq \frac{1}{(n\lambda)^n} \{R^2 Lu(x) + a_L + \Lambda\}^n, \]
where we used
\[ L\left(d_y^2/2\right) = d_y L d_y + \langle A \nabla d_y, \nabla d_y \rangle \leq a_L + \Lambda |\nabla d_y|^2 \]
in the last step to get \( L\left(d_y^2/2\right) \leq a_L + \Lambda \).

We also have to consider the other case, namely the case when \( x \) is a cut point of \( y \). In general, this kind of situation is easily overcome. Indeed, we will make use of upper barrier technique due to Calabi \([4] \); see also \([5] \). Since \( y = \exp_x \nabla (R^2u)(x) \), \( x \) is not a cut point of \( y_s = \phi_s(x) := \exp_x \nabla (sR^2u)(x) \), for all \( 0 \leq s < 1 \). By continuity, \( \text{Jac} \phi(x) = \lim_{s \to 1} \text{Jac} \phi_s(x) \). As before,

\[ \text{Jac} \phi_s(x) \leq |\text{det} D^2\left(sR^2u + d_y^2/2\right)(x)|. \]

Also,

\[ \lim_{s \to 1} |\text{det} D^2\left(sR^2u + d_y^2/2\right)(x)| = \lim_{s \to 1} |\text{det} D^2\left(R^2u + d_y^2/2\right)(x)| = \lim_{s \to 1} |\text{det} D^2w_y(x)|. \]

By the triangle inequality, we have

\[ R^2u + \frac{1}{2}|d_y + d(y, y)|^2 \geq w_y. \]

Note that the equality holds at \( x \). Since the function

\[ R^2u + \frac{1}{2}|d_y + d(y, y)|^2 = w_{y_s} + d(y_s, y)d_{y_s} + \frac{1}{2}d(y, y)^2 \]
is smooth near \( x \) and has a local minimum at \( x \) (recall that \( w_y \) has a minimum at \( x \)), its Hessian at \( x \) is nonnegative definite. Let \(-k^2 \) (\( k > 0 \)) be a lower bound of sectional curvature along the minimal geodesic joining \( x \) and \( y \). By the Hessian comparison theorem (see, e.g., \([12] \) and \([13] \)),

\[ D^2d_{y_s}(x) \leq k \coth(kd(x, y_s)) \text{Id} \leq N \text{Id}, \]
uniformly in \( s \in [\frac{1}{2}, 1] \) for some number \( N \). Therefore, for all \( s \in [\frac{1}{2}, 1] \),

\[ 0 \leq D^2w_y(x) + d(y_s, y)D^2d_{y_s}(x) \leq D^2w_y(x) + N d(y_s, y) \text{Id}. \]
In particular, \( D^2w_y(x) + N d(y_s, y) \text{Id} \) is nonnegative definite. As before, using inequality (12), we obtain \( \forall s \in [\frac{1}{2}, 1] \)

\[ \text{det} \left(D^2w_y(x) + N d(y_s, y) \text{Id}\right) \leq \frac{1}{\lambda^n} \left\{ n^{-1}Lw_{y_s}(x) + N d(y_s, y)\Lambda \right\}^n \leq \frac{1}{\lambda^n} \left\{ n^{-1} \left(R^2Lu(x) + a_L + \Lambda \right) + N d(y_s, y)\Lambda \right\}^n. \]
Since \( d(y_s, y) \to 0 \) as \( s \to 1 \), we observe
\[
0 \leq \lim_{s \to 1} \left| \det D^2 w_{y_s}(x) \right| = \lim_{s \to 1} \det \left( D^2 w_{y_s}(x) + N d(y_s, y) \text{Id} \right)
\]
\[
\leq \frac{1}{\lambda^n} \lim_{s \to 1} \left\{ n^{-1} \left( R^2 Lu(x) + a_L + \Lambda \right) + N d(y_s, y) \Lambda \right\}^n
\]
\[
= \frac{1}{(n\lambda)^n} \left\{ R^2 Lu(x) + a_L + \Lambda \right\}^n.
\]
Putting these steps together, we finally obtain
\[
\text{Jac } \phi(x) \leq \frac{1}{(n\lambda)^n} \left\{ R^2 Lu(x) + a_L + \Lambda \right\}^n \quad \forall x \in E,
\]
which proves the claim. Obviously,
\[
(13) \quad \int_E (\text{Jac } \phi) \, dV \leq \frac{1}{(n\lambda)^n} \int_E \left\{ \left( R^2 Lu(x) + a_L + \Lambda \right)^+ \right\}^n \, dV.
\]
At the beginning of the proof we have shown
\[
E \subset \{ u \leq 11/2 \} \cap B_{5R} \subset \{ u \leq 6 \} \cap B_{5R}.
\]
Now, the lemma follows from (11) and (13). □

The following technical lemma, which consists of the construction of a barrier, will be used in the proof of Lemma 3.1. It corresponds to [1, Lemma 5.5] and [3, Lemma 4.1].

**Lemma 3.3** (A barrier function). Let \( p \in M \), \( R > 0 \), and \( 0 < \delta < 1 \). There exists a continuous function \( v_\delta \) in \( B_{\delta R} = B_{\delta R}(p) \), smooth in \( M \setminus \text{Cut}(p) \) and such that:

- (a) \( v_\delta \geq 0 \) in \( B_{\delta R} \setminus B_{\delta R} \)
- (b) \( v_\delta \leq 0 \) in \( B_{2R} \)
- (c) \( R^2 L v_\delta + a_L + \Lambda \leq 0 \) a.e. in \( (B_{\delta R} \setminus B_{\delta R}) \)
- (d) \( R^2 L v_\delta \leq C_\delta \) a.e. in \( B_{\delta R} \)
- (e) \( v_\delta \geq -C_\delta \) in \( B_{\delta R} \)

for some positive constant \( C_\delta \) depending on \( \delta \).

**Proof.** We take \( v_\delta = \psi_\delta(d_p/R) \), where \( \psi_\delta \) is a smooth increasing function on \( \mathbb{R}^+ \) such that \( \psi_\delta'(0) = 0 \) and \( \psi_\delta(t) = \left( \frac{t}{\delta} \right)^{-\alpha} - \left( \frac{\delta}{t} \right)^{-\alpha} \) if \( t \geq \delta \). The number \( \alpha > 1 \) will be chosen later. Clearly, \( v_\delta \) is continuous in \( M \). Also, since \( d_p \) is smooth outside \( \text{Cut}(p) \cup \{ p \} \) and \( \psi_\delta'(0) = 0 \), \( v_\delta \) is smooth in \( M \setminus \text{Cut}(p) \).

Recall that \( \text{Cut}(p) \) is a closed set of measure zero.

Properties (a) and (b) are clear. For the rest of proof, we will denote \( \rho := d_p/R \). From the identity \( L[\varphi(u)] = \varphi'(u)Lu + \varphi''(u)\langle A\nabla u, \nabla u \rangle \), we get
\[
(14) \quad Lv_\delta = \frac{1}{R} \psi_\delta'(\rho) Ld_p + \frac{1}{R^2} \psi_\delta''(\rho) \langle A\nabla d_p, \nabla d_p \rangle
\]
\[
= \frac{1}{R^2} \rho \psi_\delta'(\rho) Ld_p + \frac{1}{R^2} \psi_\delta''(\rho) \langle A\nabla d_p, \nabla d_p \rangle \quad \text{in } M \setminus \text{Cut}(p).
\]
For $\delta \leq \rho < 5$, we have
\[
\psi'_\delta(\rho) = \frac{\alpha}{5} \left( \frac{\rho}{5} \right)^{-a-1} \quad \text{and} \quad \psi''_\delta(\rho) = -\frac{\alpha(\alpha + 1)}{5^2} \left( \frac{\rho}{5} \right)^{-a-2}.
\]
We recall that
\[
v \text{ can approximate } v \text{ in } [1, 1] \text{ which proves Property (d). Property (e) is clear.} \quad \square
\]

By Lemma 3.3. Possible existence of the cut locus $\text{Cut}(p)$ can be made less than $-(a_L + \Lambda)$ if $\alpha$ is chosen large enough. This proves Property (c).

Also, (14), (15), and the assumptions on $\psi_\delta$ imply that in $B_{5R} \setminus \text{Cut}(p)$
\[
R^2Lv_\delta \leq a_L \sup_{0<\rho<5} \frac{\psi'_\delta(\rho)}{\rho} + \Lambda \sup_{0<\rho<5} |\psi''_\delta(\rho)| < +\infty,
\]
which proves Property (d). Property (e) is clear. \quad \square

In the Proof of Lemma 3.1, we will apply Lemma 3.2 to $u + v_\delta$ with $v_\delta$ constructed in the proof of Lemma 3.3. Possible existence of the cut locus of $p$ may present some technical issue here; $u + v_\delta$ is not necessarily smooth in $B_{7R}$. To overcome this difficulty, we will again stick to Cabré’s approach in [1]. It is not hard to verify the following lemma, which asserts that one can approximate $v_\delta$ by a sequence of smooth functions; see [1, pp. 641–645].

\[\textbf{Lemma 3.4 (Cabré). Let } p \in M, \text{ } R > 0, \text{ and } \psi : \mathbb{R}^+ \to \mathbb{R} \text{ be a smooth increasing function such that } \psi'(0) = 0. \text{ Let } v = \psi \circ d_p. \text{ Then there exist a smooth function } 0 \leq \xi \leq 1 \text{ in } M \text{ with } \xi \equiv 1 \text{ in } B_{5R} := B_{5R}(p) \text{ and supp } \xi \subset B_{7R} \text{ and a sequence } \{v_k\}_{k=1}^\infty \text{ of smooth functions in } M \text{ such that}
\]

\[
\begin{cases}
    w_k \to \xi v & \text{uniformly in } M, \\
    Lw_k \to Lv & \text{a.e. in } B_{5R}, \text{ and} \\
    D^2w_k \leq C \text{ Id} & \text{in } M \text{ for some constant } C \text{ independent of } k.
\end{cases}
\]

Now, we are ready to prove Lemma 3.1.

\[\textbf{Proof of Lemma 3.1. Let } v_\delta \text{ be as in Lemma 3.3. Let } \{w_k\} \text{ be a sequence of smooth functions approximating } v_\delta \text{ in the way described in Lemma 3.4. Note that } u + v_\delta \geq 0 \text{ in } B_{7R} \setminus B_{5R} \text{ and } \inf_{B_{3R}}(u + v_\delta) \leq 1. \text{ Replacing } u + w_k \text{ by } (u + w_k + \varepsilon_k)/(1 + 2\varepsilon_k) \text{ for some sequence } 0 < \varepsilon_k \to 0 \text{ as } k \to \infty, \text{ we may assume that } u + w_k \text{ satisfies the hypothesis of Lemma 3.2, and hence (10) holds with } u \text{ replaced by } u + w_k. \text{ Let } \varepsilon > 0 \text{ and note that } \{u + w_k \leq 6\} \cap B_{5R} \subset \{u + v_\delta \leq 6 + \varepsilon\} \cap B_{5R} \text{ if } k \text{ is large enough, and that } \{(R^2L(u + w_k) + a_L + \Lambda^+)^n\} \text{ is uniformly bounded in } M \text{ independently of } k \text{ (since } D^2w_k \leq C \text{ Id). Letting } k \text{ tend to infinity and applying the dominated}
\]
convergence theorem, we get that $u + v_\delta$ satisfies (10) with $\{u \leq 6\}$ replaced by $\{u + v_\delta \leq 6 + \varepsilon\}$. Now letting $\varepsilon$ tend to zero, we obtain, since $Lu \leq f$,

$$|B_R| \leq \frac{1}{(n\lambda)^n} \int_{\{u + v_\delta \leq 6\} \cap B_{5R}} \left\{ (R^2(f + Lv_\delta) + a_L + \Lambda)^+ \right\}^n dV$$

$$= \frac{1}{(n\lambda)^n} \left[ \int_{E_1} + \int_{E_2} \left\{ (R^2(f + Lv_\delta) + a_L + \Lambda)^+ \right\}^n dV \right],$$

where $E_1 = \{u + v_\delta \leq 6\} \cap (B_{5R} \setminus B_{3R})$ and $E_2 = \{u + v_\delta \leq 6\} \cap B_{3R}$.

Using (c), (d) and (e) in Lemma 3.3 and (8), we get

$$|\partial B_R| \leq 7^n |B_R| \leq C'_\delta \left\{ R^{2n} \|f\|_{L^n(B_{7R})}^n + \left| \{u \leq M_\delta\} \cap B_{3R} \right| \right\},$$

for some constants $C'_\delta$ and $M_\delta$ depending on $\delta$. We easily conclude that (9) follows. \hfill \square

**Acknowledgment.** The author wishes to express his thank to Professor Mikhail Safonov and Jiaping Wang for valuable discussions. The author would also like to thank the anonymous referee for helpful suggestions and comments.

**References**


Received August 2, 2002 and revised January 23, 2003. The author was supported in part by NSF Grant No. DMS-9971052.

School of Mathematics
University of Minnesota
Minneapolis, MN 55455
E-mail address: skim@math.umn.edu
THE INTEGRAL KERNEL IN THE
KUZNETSOV SUM FORMULA FOR SU(n + 1, 1) (II).
THE CASE OF ONE DIMENSIONAL K-TYPES

N.P. Kisbye

The Kuznetsov sum formula relates spectral data concerning automorphic forms to geometric data concerning the intersection of a discrete subgroup with the big cell in the Bruhat decomposition. An explicit formula for the integral kernel in the Kloosterman term of this formula is given for the groups isomorphic to SU(n + 1, 1), n ≥ 2 and arbitrary one dimensional K-types.

1. Introduction.

Let $G$ be a connected semisimple Lie group of real rank one, and let $\Gamma$ be a discrete subgroup of finite covolume of $G$. There is a Kuznetsov formula in this context (see [MW]) that relates spectral data concerning automorphic forms to geometric data concerning the intersection of a discrete subgroup with the big cell in the Bruhat decomposition. The $\tau$-function is the kernel for the integral transformation relating test functions on the spectral side to those on the geometric side. In the classical case, this integral transformation can be described in terms of classical Bessel functions (see [K], [GW], [MW], Appendix), but in the general case the determination of the Kloosterman term in the Kuznetsov formula is much more complicated.

This function has been determined when $G = \text{SO}(n+1,1)$ and SU(n+1,1) (see [MW] and [Kb] respectively) in the case of the trivial $K$-type. In the present paper we shall extend the methods in [Kb] to obtain a formula for $\tau$ (see (2.7)) in the case of the group SU(n + 1, 1) and an arbitrary one dimensional $K$-type. Similarly as in [Kb] the calculation requires solving complicated recurrence relations.

Let $G = NAK$ be an Iwasawa decomposition of $G$ and let $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{k}$ be the corresponding decomposition at the Lie algebra level. Let $M$ be the centralizer of $A$ in $K$ and let $\chi$ be a nontrivial unitary character on $N$. As in [Kb], one necessary ingredient in our computations is the knowledge of generators for the $M_\chi$-invariants in the universal Lie algebra $\mathcal{U}(\mathfrak{n})$ for those groups (see [MV]). The main difference comes from the fact that in this case the recurrence formulas needed to compute the $\tau$-function involve new and complicated terms that become zero only when the $K$-type is trivial.
However, the final formula for $\tau$ in the present case ends up being similar to that obtained in [Kb], but involving a real parameter that depends on the $K$-type.

A new feature of this case is the presence of poles in the right half plane of the form $\mu - 2$, $\mu - 4$, ..., where $\mu \in \mathbb{N}$ is the $K$-type parameter.

2. Preliminaries.

We consider the Lie subalgebra of $\mathfrak{gl}(n + 2, \mathbb{C})$ given by
\[
\mathfrak{g} = \{ X \in \mathfrak{gl}(n + 2, \mathbb{C}) \mid XJ + JX^t = 0, \; \text{tr} X = 0 \}
\]
where
\[
J = \begin{pmatrix}
0 & 0 & 1 \\
0 & I_n & 0 \\
1 & 0 & 0
\end{pmatrix}
\]
and $I_n$ is the $n \times n$ identity matrix. Then $\mathfrak{g}$ is the Lie algebra $\cong SU(n+1, 1)$.

We denote by $G = SU(n+1, 1)$ the connected Lie subgroup of $\text{GL}(n+2, \mathbb{C})$ with Lie algebra $\mathfrak{g}$. Then $G$ is the Lie algebra $\cong SU(n+1, 1)$.

A Cartan involution of $\mathfrak{g}$ is given by $\theta(X) = -X^t$.

This involution induces the Cartan decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$. We take $\mathfrak{a}$ the maximal abelian subalgebra of $\mathfrak{p}$ given by $\mathfrak{a} = \mathbb{R}H$, where
\[
H = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & -1
\end{pmatrix}
\]

Let $K$ and $A$ be the connected Lie subgroups of $G$ corresponding to $\mathfrak{k}$ and $\mathfrak{a}$, respectively. Let $M$ be the centralizer of $A$ in $K$, and let $\mathfrak{m}$ be the corresponding Lie algebra of $M$. If $\alpha \in \mathfrak{a}^*$ is such that $\alpha(H) = 1$, then let $\mathfrak{n}_\alpha$ and $\mathfrak{n}_{2\alpha}$ be the root spaces associated to $\alpha$ and $2\alpha$, respectively. Let $\rho(\mathfrak{H}_0) = \frac{1}{2} \text{tr} \; \text{ad} (\mathfrak{H}_0)$, where $\mathfrak{H}_0 \in \mathfrak{a}$. Then $\rho(\mathfrak{H}) = n + 1$. We have $\mathfrak{n}_\alpha = \{ X(x) \mid x \in \mathbb{C}^n \}$ and $\mathfrak{n}_{2\alpha} = \mathbb{R}Z(i)$, where
\[
X(x) = \begin{pmatrix}
0 & x & 0 \\
0 & 0 & -ix^t \\
0 & 0 & 0
\end{pmatrix}, \quad Z(i) = \begin{pmatrix}
0 & 0 & i \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}.
\]

We also have
\[
\mathfrak{m} = \left\{ M(A) = \begin{pmatrix}
a & \ldots & 0 \\
. & A & . \\
0 & \ldots & a
\end{pmatrix} \mid A \in M_n(\mathbb{C}), A + A^t = 0, \; 2a + \text{tr} (A) = 0 \right\}.
\]

If $\mathfrak{n} = \mathfrak{n}_\alpha \oplus \mathfrak{n}_{2\alpha}$, then $\mathfrak{g}$ has the Iwasawa decomposition $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{k}$. Let $G = NAK$ be the corresponding Iwasawa decomposition at the group level. If $\bar{\mathfrak{n}} = \theta \mathfrak{n}$, the $\mathfrak{g}$ has the Iwasawa decomposition $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{k}$. Let $G = NAK$ be the corresponding Iwasawa decomposition at the group level.

If $e_1, \ldots, e_n$ denotes the canonical basis in $\mathbb{R}^n$, we set $X_j = X(e_j)$ and $X'_j = X(ie_j)$, $Y_j = -\theta X_j$, $Y'_j = -\theta X'_j$, $Z = Z(i)$, $Z' = -\theta Z$. Then
\{X_j, X'_k, \sqrt{2}Z \mid 1 \leq j, k \leq n\} is an orthonormal basis of \( n \) with respect to \( \langle , \rangle \). Note that \( [X_i, X'_j] = 2 \delta_{ij} Z \).

If \( \chi \) is a character of \( N \), then there exists \( X_\chi \in n_\alpha \) such that \( d\chi(X) = i\langle X, X_\chi \rangle = -iB(X, \theta X_\chi) \), for \( X \in n \). Set \( M_\chi = \{ m \in M \mid \text{Ad}(u)X_\chi = X_\chi \} \). As \( M \) acts transitively on the unit sphere of \( n_\alpha \) (cf. [MV, Introduction]) there is \( u_0 \in M \) such that \( \text{Ad}(u_0)X_\chi = cX_1, c \in \mathbb{R}^+ \). So \( M_\chi = u_0 M_1 u_0^{-1} \), where \( M_1 = \{ u \in M \mid \text{Ad}(u)X_1 = X_1 \} \).

The compact subgroup \( K \) of \( G \) is given by the set of matrices of the form \( k = k(\theta) = \left( \begin{array}{cc} U & 0 \\ 0 & e^{i\theta} \end{array} \right) \), with \( U \in U(n) \), \( \det(U) = e^{-i\theta} \). If \( \mu \in \mathbb{Z} \) define \( \phi_\mu(k) = \det(U)^\mu = e^{-i\mu\theta} \) and let

\[
\xi_\mu = \phi_\mu|_M.
\]

Now consider de Verma module \( M(-\nu) = U(\mathfrak{g}) \otimes_{U(\mathfrak{p})} \mathbb{C}_{-\nu-\rho} \), where \( \mathfrak{p} = \mathfrak{m} \oplus \mathfrak{a} \oplus \mathfrak{n} \) and \( \mathbb{C}_{-\nu-\rho} \) denotes the \( \mathfrak{p} \)-module \( \mathbb{C} \) with \( \mathfrak{m} \) acting by \( d\xi_\mu \), \( \mathfrak{n} \) acting by 0 and \( \mathfrak{a} \) acting by \( -\nu-\rho \), \( \nu \in \alpha_c^\ast \). Let \( M(-\nu)[\overline{\mathbb{R}}] \) denote the \( \overline{\mathbb{R}} \)-completion of \( M(-\nu) \) (see [GW], \$2\). If \( J = (j_1, j_2, \ldots, j_m) \in \mathbb{N}^m, (\mathbb{N} = \{0, 1, 2, \ldots \}) \), \( m = 2n+1 \), and \( Y(J) = Y_1^{j_1} \cdots Y_n^{j_{2n}} Z^{j_{2n}} \), then by Poincaré-Birkhoff-Witt theorem, the set \( \{ Y(J) \mid J \in \mathbb{N}^m \} \) constitutes a basis of \( U(\overline{\mathbb{R}}) \). Hence every element in \( M(-\nu)[\overline{\mathbb{R}}] \) has an expansion of the type \( \sum_I a_I Y(I) \otimes 1 \), \( a_I \in \mathbb{C} \).

A \( \chi \)-Whittaker vector is an element \( v_\chi(-\nu) \) in \( M(-\nu)[\overline{\mathbb{R}}] \) that satisfies the equation

\[
X.v = d\chi(X)v \quad \forall X \in n.
\]

Such a vector has an expression of the form

\[
v_\chi(-\nu) = \sum_{I \in \mathbb{N}^m} a_I(\chi, -\nu) Y(I) \otimes 1
\]

where the coefficients \( a_I(\chi, -\nu) \in \mathbb{C} \) are rational functions of \( \nu \). There is a unique such Whittaker vector with \( a_0(\chi, -\nu) = 1 \) (see [BM] \$6, Lemma 11, for instance). Since \( m \cdot v_\chi(-\nu) = v_\chi(-\nu) \) for \( m \in M_\chi \), we have that for each \( I \) \( Y(I) \) must be a polynomial in the \( M_\chi \)-invariants of \( U(\overline{\mathbb{R}}) \).

Let \( \chi \) be such that

\[
d\chi(X_1) = \lambda, \quad d\chi(X_i) = d\chi(X'_j) = 0 \quad i > 1, \ j \geq 1,
\]

\( \lambda \in i\mathbb{R} \). For this choice of \( \chi \), we shall use the notation \( u(\lambda, -\nu) \) for the unique \( \chi \)-Whittaker vector in \( M(-\nu)[\overline{\mathbb{R}}] \) such that \( a_0(\chi, -\nu) = 1 \). The coefficients of this \( \chi \)-Whittaker vector will be denoted \( a(\lambda, -\nu) \). Then \( u(\lambda, -\nu) \) must be an element in \( U(\overline{\mathbb{R}})^{M_1} \). We note that \( M_1 \simeq \text{SU}(n-1) \) is the subgroup of matrices in \( M \) of the form

\[
u_1(b, B) = \left( \begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & 0 & B & 0 \\
0 & B & 0 & 0 \\
0 & 0 & 0 & 0
\end{array} \right),
\]
In what follows we shall write \( \nu \) instead of \( \nu(H) \).

3. An explicit formula for the \( \chi \)-Whittaker vector.

The aim of this section is to give a formula for a Whittaker vector \( u(\lambda, \nu + \rho) \) in the \( \overline{\pi} \)-completion of the Verma module \( M(\nu + \rho) = \mathcal{U}(\mathfrak{g}) \otimes_{\mathcal{U}\mathfrak{p}} \mathbb{C}_\nu \), where \( \mathfrak{m} \) acts on \( \mathbb{C}_\nu \) by \( d\xi_\mu \) with \( \xi_\mu \) as in (2.1), \( \mathfrak{n} \) acts by 0 and \( \mathfrak{a} \) acts by \( \nu, \nu \in \mathfrak{a}^* \).

Let \( B \in U(n - 1) \). Hence \( u(\lambda, -\nu) = \sum_j Y_j f_j \), where \( f_j \) is an \( M_1 \)-invariant polynomial in \( Y_2, \ldots, Y_n, Y'_1, \ldots, Y'_n \) and \( Z' \). Then, by [MV], Theorem B, \( f_j \in \mathbb{C}[Y'_1, Z', q_1] \), where \( q_1 = \sum_{i=2}^n Y_i^2 + Y_i'^2 \).

Let \( E_{i,j} \) be the matrix in \( \mathfrak{gl}(n+2, \mathbb{C}) \) having the entry \((i,j)\) equal to 1, and all the other entries zero. In order to have simpler formulas, it is convenient to change the basis of \( \overline{\pi}_\mathbb{C} \) to

\[
\{V_1, V_2, Y_2, \ldots, Y_n, Y'_1, \ldots, Y'_n, T\},
\]

where \( V_1 = E_{2,1}, V_2 = E_{n+2,2}, T = E_{n+2,1}. \) Note that \( Y_1 = V_1 - V_2, Y'_1 = -iV_1 - iV_2, \) and \( Z' = -i\sqrt{2} T \). Let \( q \) be the element in \( \mathcal{U}(\overline{\pi})^{M_1} \),

\[
q = \sum_{i=1}^n Y_i^2 + Y_i'^2.
\]

Now it is clear that we may write

\[
(2.5) \quad u(\lambda, -\nu) = \sum_{j,k,l \geq 0} a_{j,k,l}(\lambda, -\nu) V_1^j V_2^k T_l \otimes 1 \quad n = 1
\]

\[
(2.6) \quad u(\lambda, -\nu) = \sum_{j,k,l,m \geq 0} a_{j,k,l,m}(\lambda, -\nu) V_1^j V_2^k T_l q^m \otimes 1 \quad n > 1.
\]

Now we use the \( T(\nu) \)-transform ([GW] and [MW1]) to compute the \( \tau \)-function. We see in [MW] that this \( \tau \)-function is the main ingredient in the Bessel transform in the Kloosterman term of the sum formula of Kuznetsov type and also it appears in the Fourier coefficients of the Poincaré series. To give the formula of the \( \tau \)-function we consider two parabolic \( \Gamma \)-percuspidal subgroups \( P \) and \( P' \). Then \( P = \Gamma M \) and \( P' = \Gamma' A \Gamma' M' \). Let \( \chi \) and \( \chi_1 \) be nontrivial unitary characters on \( N \) and \( N' \) respectively. If \( W(A) = \{1, s\} \) is the Weyl group of \( (P, A) \) then we take \( s^* \) a representative of \( s \) in \( K \). The \( \tau \)-function is given by the following formula (see [MW] Proposition 1.2):

\[
(2.7) \quad \tau(\chi_1, \chi, ua, \nu) = \sum_{I \in \mathbb{N}^m} a_I(\chi, \nu) d\chi_1(Ad(uas^*))^{-1} Y(I)^T \otimes 1
\]

\( u \in M \), \( a \in A \) and where the coefficients \( a_I \) are given by Formula (2.3). Here, \( Y \mapsto Y^T \) is the automorphism of the universal enveloping algebra given by \( X \mapsto -X \), for \( X \in \mathfrak{g} \).
Lemma 3.1. Let \( X_1, X_1', V_1, V_2, T \) and \( H \) be as above. Let \( U^+ = \frac{1}{2}[H + iM(2iE_{11})] \) and \( U^- = \frac{1}{2}[-H + iM(2iE_{11})] \). Then, the following commutation relations hold for \( j, k, l \geq 1 \):

\[
[X_1, V_1^j] = jV_1^{j-1}(U^+ - j + 1)
\]
\[
[X_1, V_2^k] = kV_2^{k-1}(U^- + k - 1)
\]
\[
[X_1, T^l] = -i(V_1 + V_2)T^{l-1}
\]
\[
[X_1', V_1^j] = jV_1^{j-1}(iU^+ - i(j - 1))
\]
\[
[X_1', V_2^k] = kV_2^{k-1}(-iU^- - i(k - 1))
\]
\[
[X_1', T^l] = il(V_1 - V_2)T^{l-1}.
\]

Also we have: \( [U^+, V_2] = V_2, [U^+, T] = -T \) and \( [U^-, T] = T \).

Proof. The formulas are proved by the usual \( \text{Sl}(2, \mathbb{C}) \)-technique. They are based on the identities: \( [X_1, V_1] = U^+, [X_1, V_2] = U^-, [X_1', V_1] = iU^+, [X_1', V_2] = -iU^- \) and on the fact that \( [V_1, V_2] = -T \). \( \square \)

Thus, it follows from Lemma 3.1 that:

\[
(3.1) \quad X_1V_1^j V_2^k T^l \otimes 1
\]
\[
= (X_1V_1^j)V_2^k T^l \otimes 1 + V_1^j(X_1V_2^k)T^l \otimes 1 + V_1^jV_2^k (X_1T^l) \otimes 1
\]
\[
= jV_1^{j-1}(U^+ - j + 1)V_2^k T^l \otimes 1 + kV_1^jV_2^{k-1}(U^- + k - 1)T^l \otimes 1
\]
\[
- iV_1^jV_2^k (V_1 + V_2)T^{l-1} \otimes 1
\]
\[
= jV_1^{j-1}V_2^k T^l (k - l - j + 1 + U^+) \otimes 1
\]
\[
+ kV_1^jV_2^{k-1} T^l (l - k + 1 + U^-) \otimes 1
\]
\[
- iV_1^jV_2^k V_1 T^{l-1} \otimes 1 - iV_1^jV_2^k T^{l-1} \otimes 1.
\]

Using that \( [V_1, V_2^k] = -kV_2^{k-1}T \) and the fact that \( U^+ \) and \( U^- \) act on \( \mathbb{C}_\nu \) by \( \frac{\nu + \mu}{2} \) and \( -\frac{\nu + \mu}{2} \) respectively, the last expression is equal to

\[
(3.2) \quad jV_1^{j-1}V_2^k T^l \left( k - l - j + 1 + \frac{\nu + \mu}{2} \right) \otimes 1
\]
\[
+ kV_1^jV_2^{k-1} T^l \left( k - 1 - \frac{\nu - \mu}{2} \right) \otimes 1
\]
\[
- iV_1^jV_2^k T^{l-1} \otimes 1 - iV_1^jV_2^k T^{l-1} \otimes 1.
\]
Now, Formula (2.2) implies that:

\[
\sum_{j,k,l \geq 0} a_{j,k,l}(\lambda, \nu + \rho) X_1 V_1^j V_2^k T^l \otimes 1 \\
= \lambda \sum_{j,k,l \geq 0} a_{j,k,l}(\lambda, \nu + \rho) V_1^j V_2^k T^l \otimes 1,
\]

\[
\sum_{j,k,l \geq 0} a_{j,k,l}(\lambda, \nu + \rho) X_1 V_1^j V_2^k T^l \otimes 1 = 0.
\]

From Formulas (3.2) and (3.3) we obtain the following recurrence relation for the coefficients \(a_{j,k,l}(\lambda, \nu + \rho)\):

\[
\lambda a_{j,k,l}(\lambda, \nu + \rho) \\
= (j + 1) \left( k - l - j + \frac{\nu + \mu}{2} \right) a_{j+1,k,l}(\lambda, \nu + \rho) \\
+ (k + 1) \left( k - \frac{\nu - \mu}{2} \right) a_{j,k+1,l}(\lambda, \nu + \rho) \\
- (l + 1) a_{j-1,k,l+1}(\lambda, \nu + \rho) - (l + 1) a_{j,k,l+1}(\lambda, \nu + \rho).
\]

Also, we have

\[
X_1' V_1^j V_2^k T^l \otimes 1 \\
= (X_1' V_1) V_2^k T^l \otimes 1 + V_1^j (X_1' V_2^k) T^l \otimes 1 + V_1^j V_2^k (X_1', T^l) \otimes 1 \\
= i j V_1^j V_2^k k (U^+ - j + 1 + k - l) \otimes 1 \\
+ V_1^j V_2^k (i l (V_1 - V_2)) T^{l-1} \otimes 1 \\
= i j V_1^j V_2^k k (U^+ - j + 1) \otimes 1 \\
+ V_1^j V_2^k k (U^+ - i (k - 1 + l)) \otimes 1 \\
+ i l V_1^j V_2^k V_1 T^{l-1} \otimes 1 - i l V_1^j V_2^k V_1 T^{l-1} \otimes 1 \\
= i j V_1^j V_2^k k (U^+ - j + 1 + k - l + \frac{\nu + \mu}{2}) \otimes 1 \\
+ i V_1^j V_2^k k (U^+ - k + 1 + \frac{\nu - \mu}{2}) \otimes 1 \\
+ i V_1^j V_2^k k (U^+ - i (k - 1 + l)) \otimes 1 \\
+ i l V_1^j V_2^k V_1 T^{l-1} \otimes 1 - i l V_1^j V_2^k V_1 T^{l-1} \otimes 1.
\]
This computation together with Formula (3.4) gives a second recurrence relation for the coefficients $a_{j,k,l}(\lambda, \nu + \rho)$:

\begin{equation}
(j + 1) \left( -j + k - l + \frac{\nu + \mu}{2} \right) a_{j+1,k,l}(\lambda, \nu + \rho) \\
+ (k + 1) \left( -k + \frac{\nu - \mu}{2} \right) a_{j,k+1,l}(\lambda, \nu + \rho) \\
+ (l + 1) a_{j-1,k,l+1}(\lambda, \nu + \rho) - (l + 1) a_{j,k-1,l+1}(\lambda, \nu + \rho) = 0.
\end{equation}

Finally, from the identities (3.5) and (3.7) we get the following relations:

\begin{align}
\lambda a_{j,k,l}(\lambda, \nu + \rho) &= (j + 1)(\nu + \mu - 2j + 2k - 2l)a_{j+1,k,l}(\lambda, \nu + \rho) \\
&
- 2(l + 1)a_{j,k-1,l+1}(\lambda, \nu + \rho),
\end{align}

(3.8)

\begin{align}
\lambda a_{j,k,l}(\lambda, \nu + \rho) &= (k + 1)(-\nu + \mu + 2k)a_{j,k+1,l}(\lambda, \nu + \rho) \\
&
- 2(l + 1)a_{j-1,k,l+1}(\lambda, \nu + \rho).
\end{align}

(3.9)

**Proposition 3.2.** Let $\chi$ be as in (2.4). If $u(\lambda, \nu + \rho)$ denotes the canonical $\chi$-Whittaker vector in $M(\nu + \rho)$ for $SU(2, 1)$, then the coefficients $a_{j,k,l}(\lambda, \nu + \rho)$ of $u(\lambda, \nu + \rho)$ are given by the formula:

\begin{equation}
a_{j,k,l}(\lambda, \nu + \rho) = \frac{\chi^{j+k+2l}(-1)^{j+k+l-1}}{2l!j!!!(\nu + \mu - 2i) \prod_{i=0}^{k-1} (\nu - \mu - 2i) \prod_{i=0}^{j+l-1} (\nu + 1 - i)} \prod_{i=k+l}^{j+k+l-1} (\nu + 1 - i)
\end{equation}

(3.10)

$a_{0,0,0}(\lambda, \nu + \rho) = 1$.

The proof of this proposition follows by induction on $n = j + k + 2l$ and is very similar to the proof of Proposition 3.2 in $[\text{Kb}]$.

Now we shall compute the coefficients of the $\chi$-Whittaker vector for $SU(n+1, 1)$ in the case $n > 1$. We set

\begin{equation}
M_{jk} = M(E_{jk} - E_{kj}), \quad M^{jk} = M(iE_{jk} + iE_{kj}).
\end{equation}

(3.11)

Then $d\xi(M_{ij}) = 0$ and $d\xi(M^{ij}) = \delta_{ij}(-i\mu)$. 
Lemma 3.3. Let $X_i, X'_i, Y_i, Y'_i, V_1, V_2, T$ and $q$ be as in Section 1. Then the following identities hold: For $i \geq 2$,

$$[X_i, V_1^j] = jV_1^{j-1}M_i \quad \text{where} \quad M_i = \frac{1}{2}(M(E_{i1} - E_{1i}) + iM(iE_{i1} + iE_{1i}))$$

$$[X_i, V_2^k] = kV_2^{k-1}(M_i)^t$$

$$[X_i, T^t] = -iT^{t-1}(Y_i')$$

$$[X'_i, V_1^j] = ijV_1^{j-1}M_i$$

$$[X'_i, V_2^k] = -ikV_2^{k-1}(M_i)^t$$

$$[X'_i, T^t] = iT^{t-1}(Y_i')$$

$$[M_i, V_2^k] = (k/2)(-Y_i + iY'_i)V_2^{k-1}.$$  

For $i \geq 1$,

(3.12) \hspace{1em} \[ [X_i, q^m] = 2 \sum_{s=0}^{m-1} 4^s U_i^{(s)} T^s q^{m-1-s} \sum_{j=s}^{m-1} \binom{j}{s} (n + H - 2(m - 1 - j)) \]

$$+ 2 \sum_{s=0}^{m-1} 4^s \left( \binom{m}{s+1} T^s q^{m-1-s} \Phi(s)(X_i) \right),$$

(3.13) \hspace{1em} \[ [X'_i, q^m] = 2i \sum_{s=0}^{m-1} 4^s U_i^{(s-1)} T^s q^{m-1-s} \sum_{j=s}^{m-1} \binom{j}{s} (n + H - 2(m - 1 - j)) \]

$$- 2 \sum_{s=0}^{m-1} 4^s \left( \binom{m}{s+1} T^s q^{m-1-s} \Phi(s)(X'_i) \right),$$

where $U_i^{(s)} = Y_k$ if $s$ is even and $-iY'_k$ if $s$ is odd. $\Phi(s)(X_i)$ and $\Phi(s)(X'_i)$ belong to $U(g)m$ and are given by the formulas

$$\Phi(s)(X_i) = \begin{cases} 
\sum_{j=1}^{n} (Y_j M_{ij} + Y'_j M^{ij}), & \text{if } s \text{ is even}, \\
-\sum_{j=1}^{n} (Y'_j M_{ij} - Y_j M^{ij}), & \text{if } s \text{ is odd}.
\end{cases}$$

and

$$\Phi(s)(X'_i) = \begin{cases} 
\sum_{j=1}^{n} (Y_j M^{ij} + Y'_j M_{ij}), & \text{if } s \text{ is even}, \\
\sum_{j=1}^{n} (Y'_j M^{ij} - Y_j M_{ij}), & \text{if } s \text{ is odd}.
\end{cases}$$
Proof. For the first formulas see [Kb], Lemma 3.1. To prove Formula (3.12) we proceed by induction on \( m \). The proof of (3.13) is similar. We need the formulas:

\[
\begin{align*}
[[X_i, Y_j], Y_k] &= -\delta_{ij} Y_k + \delta_{jk} Y_i - \delta_{ik} Y_j \\
[[X_i, Y_j'], Y_k'] &= \delta_{ij} Y_k + \delta_{jk} Y_i + \delta_{ik} Y_j \\
[X_i, Y_j] &= \delta_{ij} H + M(E_{ij} - E_{ji}) \\
[X_i, Y_j'] &= M(i(E_{ij} + E_{ji})) \\
[X_i', Y_j] &= -M(i(E_{ij} + E_{ji})) \\
[X_i', Y_j'] &= \delta_{ij} H + M(E_{ij} - E_{ji}) \\
[[X_i', Y_j], Y_k] &= \delta_{ij} Y_k' + \delta_{jk} Y_i' + \delta_{ik} Y_j' \\
[[X_i', Y_j'], Y_j'] &= -\delta_{ij} Y_k' + \delta_{jk} Y_i' - \delta_{ik} Y_j'.
\end{align*}
\]

For \( m = 1 \) we have

\[
[X_i, q] = \sum_{j=1}^{n} ([X_i, Y_j^2] + [X_i, Y_j'^2])
\]

\[
= \sum_{j=1}^{n} ([[[X_i, Y_j], Y_j] + [[[X_i, Y_j'], Y_j'] + 2Y_j[X_i, Y_j] + 2Y_j'[X_i, Y_j']])
\]

\[
= \sum_{j=1}^{n} (-2\delta_{ij} Y_j + Y_i + 2\delta_{ij} Y_j + Y_i) + 2Y_i H + 2Y_j M_{ij} + 2Y_j' M_{ij}
\]

\[
= 2Y_i (n + H) + 2 \sum_{j=1}^{n} (Y_j M_{ij} + Y_j' M_{ij})
\]

\[
= 2Y_i (n + H) + 2\Phi^{(0)}(X_i).
\]

If \( m > 1 \) we need the following identities:

\[
\sum_{j=s}^{m-1} \binom{j}{s} (n + H - 2(m - 1 - j)) q = q \sum_{j=s}^{m-1} \binom{j}{s} (n + H - 2(m - j)),
\]

(3.16)

\[
[q^m, Y_i] = \sum_{s=1}^{m} 4^s U_i^{(s)} T^s q^{m-s},
\]

(3.17)

\[
[\Phi^{(s)}, q] = 4T\Phi^{(s+1)}.
\]

Formula (3.15) follows from the fact that \([H, q] = -2q\), and (3.16) was proved in [Kb], Lemma 3.3. To prove (3.17) we observe that if \( s \) is even
then

\[
[\Phi^{(s)}(X_i), q] = \sum_{j=1}^{n} \left( [Y_j, q] M_{ij} + [Y'_j, q] M^{ij} \right)
= 4iT \sum_{j=1}^{n} (Y'_j M_{ij} - Y_j M^{ij}) = 4T \Phi^{(s+1)}(X_i).
\]

If \( s \) is odd then

\[
[\Phi^{(s)}(X_i), q] = i \sum_{j=1}^{n} \left( [Y'_j, q] M_{ij} - [Y_j, q] M^{ij} \right)
= 4T \sum_{j=1}^{n} (Y'_j M_{ij} + Y_j M^{ij}) = 4T \Phi^{(s+1)}(X_i).
\]

Thus (3.17) follows.

Using that \([X_i, q^{m+1}] = [X_i, q^m]q + q^m[X_i, q]\) we have

\[
[X_i, q^{m+1}] = 2 \sum_{s=0}^{m-1} 4^s U_i^{(s)} T^s q^{m-s} \sum_{j=s}^{m-1} \binom{j}{s} (n + H - 2(m - j))
+ 2q^m Y_i(n + H) + 2 \sum_{s=0}^{m-1} 4^s \left( \binom{m}{s+1} T^s q^{m-s} \Phi^{(s)}(X_i) \right)
+ 2 \sum_{s=0}^{m-1} 4^{s+1} \binom{m}{s+1} T^{s+1} q^{m-(s+1)} \Phi^{(s+1)}(X_i) + 2q^m \Phi^{(0)}(X_i),
\]

thus

(3.18)

\[
[X_i, q^{m+1}]
= 2 \sum_{s=0}^{m-1} 4^s U_i^{(s)} T^s q^{m-s} \sum_{j=s}^{m-1} \binom{j}{s} (n + H - 2(m - j))
+ 2 U_i^{(0)} q^m (n + H) + 2 \sum_{s=1}^{m} 4^s \left( \binom{m}{s} T^s q^{m-s} (n + H) \right)
+ 2 \left( \binom{m}{1} q^m \Phi^{(0)}(X_i) + 2 \sum_{s=1}^{m-1} 4^s \left( \binom{m}{s+1} + \binom{m}{s} \right) T^s q^{m-s} \Phi^{(s)}(X_i) \right)
+ 2 4^m \binom{m}{m} T^m \Phi^{(m)}(X_i) + 2q^m \Phi^{(0)}(X_i).
\]
Now,
\begin{equation}
2U_i^{(0)}q^m \sum_{j=0}^{m-1} \binom{j}{0} (n + H - 2(m - j)) + 2U_i^{(0)}q^m (n + H)
\end{equation}
and
\begin{equation}
2 \sum_{s=1}^{m} 4^s U_i^{(s)} T^s q^{m-s} \sum_{j=s}^{m-1} \binom{j}{s} (n + H - 2(m - j))
\end{equation}

Using the identity \( \binom{m}{s+1} + \binom{m}{s} = \binom{m+1}{s+1} \) we have that the sum of the last four terms in Formula (3.18) equals
\[ 2 \sum_{s=0}^{m} 4^s \binom{m+1}{s+1} T^s q^{m-s} \Phi(s)(X_i). \]

Thus we obtain
\begin{equation}
[X_i, q^{m+1}] = 2 \sum_{s=0}^{m} 4^s U_i^{(s)} T^s q^{m-s} \sum_{j=s}^{m} \binom{j}{s} (n + H - 2(m - j))
\end{equation}
so the proof of Formula (3.12) is complete. Using similar computations and the identity
\[ [q^m, Y'_i] = \sum_{s=1}^{m} 4^s \binom{m-1}{s} U_i^{(s-1)} T^s q^{m-s}, \]
whose proof can also be found in [Kb], Lemma 3.3, one can obtain Formula (3.13).
Now, it follows that for \( i \geq 2 \)

\[
(3.22) \quad X_i \cdot V_i^j V_2^k T^l q^m \otimes 1 \\
= j V_i^{j-1} M_i V_2^k T^l q^m \otimes 1 \\
+ k V_i^j V_2^{k-1} M_i T^l q^m \otimes 1 - il V_i^j V_2^k T^{l-1} Y_i' q^m \otimes 1 \\
+ V_i^j V_2^k T^l X_i \cdot q^m \otimes 1.
\]

As \( m \) acts by \( d\xi_\mu \) and from the definition of \( M_{kl} \) and \( M^{kl} \) in (3.11), we have that \( M_i \otimes 1 = M_i^l \otimes 1 = 0 \) for \( i \geq 2 \) and

\[
\Phi(s)(X_i) \otimes 1 = \begin{cases} 
Y_i' M^{ii} \otimes 1 = -i\mu Y_i', & \text{if } s \text{ is even} \\
-iY_i M^{ii} \otimes 1 = -\mu Y_i, & \text{if } s \text{ is odd.}
\end{cases}
\]

Using that \([m, T] = [m, q] = 0\) and setting \( \sigma(m - 1, s, \nu) = \sum_{j=s}^{m-1} (j)(n + \nu - 2(m - 1 - j)) \) it follows that

\[
(3.23) \\
X_i \cdot V_i^j V_2^k T^l q^m \otimes 1 \\
= \frac{jk}{2} V_i^{j-1} (-Y_i + iY_i') V_2^k T^{l-1} q^m \otimes 1 - il V_i^j V_2^k T^{l-1} Y_i' q^m \otimes 1 \\
+ 2Y_i \sum_{s=0, s \text{ even}}^{m-1} 4^s \sigma(m - 1, s, \nu) V_i^j V_2^k T^{l+s} q^{m-1-s} \\
- 2iY_i' \sum_{s=0, s \text{ odd}}^{m-1} 4^s \sigma(m - 1, s, \nu) V_i^j V_2^k T^{l+s} q^{m-1-s} \\
- 2i\mu Y_i' \sum_{s=0, s \text{ even}}^{m-1} 4^s \left( \begin{array}{c}
m \\s+1\end{array} \right) V_i^j V_2^k T^{l+s} q^{m-1-s} \\
- 2\mu Y_i \sum_{s=0, s \text{ odd}}^{m-1} 4^s \left( \begin{array}{c}
m \\s+1\end{array} \right) V_i^j V_2^k T^{l+s} q^{m-1-s} \\
= Y_i \left( -\frac{jk}{2} V_i^{j-1} V_2^k T^{l+1} q^m \otimes 1 + 2 \sum_{s=0, s \text{ even}}^{m-1} 4^s \sigma(m - 1, s, \nu) V_i^j V_2^k T^{l+s} q^{m-1-s} \\
- 2\mu \sum_{s=0, s \text{ odd}}^{m-1} 4^s \left( \begin{array}{c}
m \\s+1\end{array} \right) V_i^j V_2^k T^{l+s} q^{m-1-s} \right) \\
- iY_i' \left( -\frac{jk}{2} V_i^{j-1} V_2^k T^{l+1} q^m \otimes 1 + il V_i^j V_2^k T^{l-1} q^m \otimes 1 \right).
\]
ON THE INTEGRAL KERNEL FOR KUZNETSOV SUM FORMULA 307

\[ + 2 \sum_{s=0, s \text{ odd}}^{m-1} 4^s \sigma(m - 1, s, \nu)V_1^j V_2^{kT^l+s} q^{m-1-s} \]

\[ \cdot \sum_{s=0, s \text{ even}}^{m-1} 4^s \left( \frac{m}{s+1} \right) V_1^j V_2^{kT^l+s} q^{m-1-s} \).

Using that \( X_i \cdot u(\lambda, \nu) = 0 \) for \( i \geq 2 \) and the linear independence of the monomials \( Y_i V_1^j V_2^{kT^l} q^m \) and \( Y'_i V_1^j V_2^{kT^l} q^m \) we obtain from the last identity two recurrence formulas for the coefficients \( a_{j,k,l,m} \):

\[ (j+1)(k+1)a_{j+1,k+1,l,m} \]

\[ = 2 \sum_{s=0}^{l} 4^s \sigma(m + s, s, \nu)a_{j,k,l-s,m+1+s} \]

\[ + (l+1)a_{j,k,l+1,m} + 2\mu \sum_{s=0}^{l} (-1)^s \left( \frac{m + 1 + s}{s + 1} \right) a_{j,k,l-s,m+1+s}. \]

\[ (l+1)a_{j,k,l+1,m} = 2 \sum_{s=0}^{l} (-1)^s 4^s \sigma(m + s, s, \nu)a_{j,k,l-s,m+1+s} \]

\[ - 2\mu \sum_{s=0}^{l} 4^s \left( \frac{m + s + 1}{s + 1} \right) a_{j,k,l-s,m+1+s}. \]

To compute the action of \( X_1 \) we need the identities

\[ V_2^k Y_1 = k V_2^{k-1} T + V_1 V_2^k - V_2^{k+1} \]

\[ V_2^k Y_1' = -i k V_2^{k-1} T + i V_1 V_2^k - i V_2^{k+1} \]

and the commutation relations in Lemma 3.1. Thus

\[ (3.26) \quad X_1 \cdot V_1^j V_2^{kT^l} q^m \otimes 1 \]

\[ = j \left( \frac{\nu + \mu}{2} + k - l - j - m + 1 \right) V_1^{j-1} V_2^{kT^l} q^m \otimes 1 \]

\[ + k \left( \frac{\mu - \nu}{2} + k + m - 1 \right) V_1^j V_2^{k-1T^l} q^m \otimes 1 \]

\[ - iV_1^{j+1} V_2^{kT^l-1} q^m \otimes 1 - iV_1^j V_2^{k+1T^{l-1}} q^m \otimes 1 \]

\[ + 2 \sum_{s=0, s \text{ even}}^{m-1} 4^s \sigma(m - 1, s, \nu) \left( k V_1^j V_2^{k-1T^l+s+1} q^{m-1-s} \right. \]

\[ + V_1^{j+1} V_2^{kT^l+s} q^{m-1-s} - V_1^j V_2^{k+1T^l+s} q^{m-1-s} \)
Therefore we get

\[
- 2 \sum_{s=0, s \text{ odd}} 4^s \sigma(m - 1, s, \nu) \left( kV_1^j V_2^{k-1} T^{l+s+1} q^{m-1-s} + V_1^j V_2^{k+1} T^{l+s+1} q^{m-1-s} \right) \\
+ V_1^{j+1} V_2^{k-1} T^{l+s} q^{m-1-s} + V_1^{j+1} V_2^{k+1} T^{l+s} q^{m-1-s} \\
- 2\mu \sum_{s=0, s \text{ even}} 4^s \left( \frac{m}{s+1} \right) \left( kV_1^j V_2^{k-1} T^{l+s+1} q^{m-1-s} + V_1^{j+1} V_2^{k-1} T^{l+s+1} q^{m-1-s} + V_1^j V_2^{k+1} T^{l+s+1} q^{m-1-s} \right) \\
+ V_1^{j+1} V_2^{k-1} T^{l+s} q^{m-1-s} - V_1^{j+1} V_2^{k+1} T^{l+s} q^{m-1-s}. 
\]

To compute the action of $X'_1$ we shall use that

\[
\Phi^{(s)}(X'_1) = \begin{cases} 
Y_1 M^{11} \otimes 1 = -i\mu Y_1, & \text{if } s \text{ is even}, \\
iY_1 M^{11} \otimes 1 = \mu Y'_1, & \text{if } s \text{ is odd}.
\end{cases}
\]

Therefore we get

\[
X'_1 \cdot V_1^j V_2^k T^l q^m \otimes 1 \\
= ij \left( \frac{\nu + \mu}{2} + k - l - j - m + 1 \right) V_1^{j-1} V_2^{k} T^l q^m \otimes 1 \\
- ik \left( \frac{\mu - \nu}{2} + k + m - 1 \right) V_1^j V_2^{k-1} T^l q^m \otimes 1 \\
+ ilV_1^{j+1} V_2^{k-1} T^l q^m \otimes 1 - ilV_1^j V_2^{k+1} T^l q^m \otimes 1 \\
- 2i \sum_{s=0, s \text{ even}} 4^s \sigma(m - 1, s, \nu) \left( kV_1^j V_2^{k-1} T^{l+s+1} q^{m-1-s} + V_1^j V_2^{k+1} T^{l+s+1} q^{m-1-s} \right) \\
+ V_1^{j+1} V_2^{k-1} T^{l+s} q^{m-1-s} + V_1^{j+1} V_2^{k+1} T^{l+s} q^{m-1-s} \\
+ 2i \sum_{s=0, s \text{ odd}} 4^s \sigma(m - 1, s, \nu) \left( kV_1^j V_2^{k-1} T^{l+s+1} q^{m-1-s} + V_1^j V_2^{k+1} T^{l+s+1} q^{m-1-s} \right) \\
+ V_1^{j+1} V_2^{k-1} T^{l+s} q^{m-1-s} - V_1^{j+1} V_2^{k+1} T^{l+s} q^{m-1-s} \\
+ 2i\mu \sum_{s=0, s \text{ even}} 4^s \left( \frac{m}{s+1} \right) \left( kV_1^j V_2^{k-1} T^{l+s+1} q^{m-1-s} + V_1^j V_2^{k+1} T^{l+s+1} q^{m-1-s} \right) \\
+ V_1^{j+1} V_2^{k-1} T^{l+s} q^{m-1-s} - V_1^{j+1} V_2^{k+1} T^{l+s} q^{m-1-s} \\
+ 2i\mu \sum_{s=0, s \text{ odd}} 4^s \left( \frac{m}{s+1} \right) \left( kV_1^j V_2^{k-1} T^{l+s+1} q^{m-1-s} + V_1^j V_2^{k+1} T^{l+s+1} q^{m-1-s} \right).
\]
\[ + V_1^{j+1}V_2^{k+1}T^{l+s}q^{m-1-s} + V_1^jV_2^{k+1}T^{l+s}q^{m-1-s} \].

Thus

(3.28)
\[
(X_1 - iX'_1) \cdot V_1^j V_2^k T^l q^m \otimes 1
= j(\nu + \mu + 2k - 2l - 2j - 2m + 2)V_1^{j-1}V_2^k T^l q^m \otimes 1
- 2V_1^j V_2^{k+1} T^{l-1} q^m \otimes 1 - 4 \sum_{s=0}^{m-1} 4^s \sigma(m - 1, s, \nu)V_1^j V_2^{k+1} T^{l+s} q^{m-1-s}
- 4\mu \sum_{s=0}^{m-1} (-1)^s 4^s \left( \binom{m}{s+1} V_1^j V_2^{k+1} T^{l+s} q^{m-1-s} \otimes 1, \right.
\]

and

(3.29)
\[
(X_1 + iX'_1) \cdot V_1^j V_2^k T^l q^m \otimes 1 =
k(\mu - \nu + 2k + 2m - 2)V_1^j V_2^{k-1} T^{l} q^m \otimes 1 - 2V_1^{j+1} V_2^k T^{l} q^m \otimes 1
+ 4k \sum_{s=0}^{m-1} (-1)^s 4^s \sigma(m - 1, s, \nu)V_1^j V_2^{k-1} T^{l+s+1} q^{m-1-s}
+ 4 \sum_{s=0}^{m-1} (-1)^s 4^s \sigma(m - 1, s, \nu)V_1^{j+1} V_2^k T^{l+s} q^{m-1-s}
- 4\mu k \sum_{s=0}^{m-1} 4^s \left( \binom{m}{s+1} V_1^j V_2^{k-1} T^{l+s+1} q^{m-1-s} \otimes 1, \right.
\]

Proposition 3.4. The coefficients \( a_{j,k,l,m}(\nu+\rho, \lambda) \) for the \( \chi \)-Whittaker \( u(\lambda, \nu + \rho) \) in \( M(\nu + \rho) \) satisfy the following recurrence relations:

(3.30)
\[
\lambda a_{j,k,l,m} = (j + 1)(\nu + \mu + 2k - 2l - 2m)a_{j+1,k,l,m} - 2(l + 1)a_{j,k-1,l+1,m}
- 4 \sum_{s=0}^{l} 4^s \left( \sigma(m + s, s, \nu) + \mu(-1)^s \left( \binom{m + 1 + s}{s + 1} \right) a_{j,k-1,l-s,m+1+s}, \right.
\]
\[
\begin{align*}
(3.31) & \quad \lambda a_{j,k,l,m} = (k + 1)(\mu - \nu + 2k + 2m)a_{j,k+1,l,m} - 2(l + 1)a_{j-1,k,l+1,m} \\
& \quad + 4(k + 1) \sum_{s=0}^{l-1} 4^s \left( (-1)^s (m + s, s, \nu) - \mu \binom{m + 1 + s}{s + 1} \right) a_{j,k,l-s-1,m+1+s} \\
& \quad + 4 \sum_{s=0}^{l} 4^s \left( (-1)^s (m + s, s, \nu) - \mu \binom{m + 1 + s}{s + 1} \right) a_{j-1,k,l-s,m+1+s}
\end{align*}
\]

\[
(3.32) \quad (j + 1)(k + 1)a_{j+1,k+1,l,m} = (l + 1)a_{j,k,l+1,m} \\
+ 2 \sum_{s=0}^{l} 4^s \left( \sigma(m + s, s, \nu) + \mu(-1)^s \binom{m + 1 + s}{s + 1} \right) a_{j,k,l-s,m+1+s}
\]

\[
(3.33) \quad (l + 1)a_{j,k,l+1,m} = 2 \sum_{s=0}^{l} 4^s \left( (-1)^s \sigma(m + s, s, \nu) \right. \\
- \mu \binom{m + s + 1}{s + 1} \left. \right) a_{j,k,l-s,m+1+s}.
\]

Proof. The equations follow from the definition of the Whittaker vector, Formulas (3.24), (3.25), (3.28) and (3.29). We note that if we set \( \mu = 0 \) we get the formulas given in \([Kb]\), Proposition 3.5. \( \square \)

Equation (3.33) implies that \( a_{j,k,l,m} = b_{j,k,l,m}a_{j,k,0,l+m} \), where \( b_{j,k,l,m} \) satisfies the formulas:

\[
(3.34) \quad b_{j,k,0,m} = 1
\]

\[
(3.35) \quad (l + 1)b_{j,k,l,m} = 2 \sum_{s=0}^{l} 4^s \left( (-1)^s \sigma(m + s, s, \nu) \right. \\
- \mu \binom{m + s + 1}{s + 1} \left. \right) b_{j,k,l-s,m+1+s}.
\]

From (3.34) and (3.35) we conclude that \( b_{j,k,l,m} \) doesn’t depend on \( j \) and \( k \). Therefore we shall use the notation \( b_{l,m} \) instead of \( b_{j,k,l,m} \).

Formulas (3.30) and (3.32) imply that

\[
(3.36) \quad \lambda a_{j,k,l,m} = (j + 1)(\nu + \mu - 2l - 2m)a_{j+1,k,l,m},
\]

hence

\[
(3.37) \quad a_{j,k,l,m} = \frac{\lambda^j}{j! \prod_{i=l+m}^{j+l+m-1} (\nu + \mu - 2i)} a_{0,k,l,m}.
\]
From Formulas (3.31) and (3.33) we get
\[
\lambda a_{j,k,l,m} = (k + 1)(\mu - \nu + 2k + 2m + 2l)a_{j,k,l,m},
\]
which implies
\[
a_{j,k,l,m} = \frac{(-1)^k \lambda^k}{k! \prod_{i=l+m} (\nu - \mu - 2i)} a_{j,0,l,m}.
\]

Finally, (3.37) together with (3.39) implies
\[
a_{j,k,l,m} = \frac{(-1)^k \lambda^j + k}{j! k! \prod_{i=l+m} (\nu + \mu - 2i) \prod_{i=k} (\nu - \mu - 2i)} a_{0,0,l,m}.
\]

If we set \( l = 0 \) in the Formulas (3.32) and (3.33) we get:
\[
(j + 1)(k + 1)a_{j+1,k+1,0,m} = 4\sigma(m, 0, \nu)a_{j,k,0,m+1}.
\]

But \( \sigma(m, 0, \nu) = \sum_{j=0}^{m} (n + \nu - 2(m - j)) = (m + 1)(n + \nu - m) \) and this together with (3.36) and (3.38) imply
\[
a_{j,k,0,m+1} = -\lambda^2 4(m + 1)(n + \nu - m)(\nu + \mu - 2m)(\nu - \mu - 2k - 2m)a_{j,k,0,m}
\]
and then
\[
a_{j,k,0,m} = \frac{(-1)^m \lambda^{2m}}{4^m m! \prod_{i=0}^{m-1} (n + \nu - i) \prod_{i=j}^{k} (\nu + \mu - 2i) \prod_{i=0}^{k-1} (\nu - \mu - 2i)} a_{j,k,0,0}.
\]

Finally, we obtain the following explicit formula for the coefficient \( a_{j,k,l,m} \):
\[
a_{j,k,l,m}(\nu, \lambda) = b_{l,m}(\nu) a_{j,k,0,l+m}(\nu, \lambda)
\]
\[
= \frac{b_{l,m}(\nu)(-1)^{j+l+m}\lambda^{j+k+l+2m}}{j!k!(l+m)!4^{l+m} \prod_{i=0}^{j} (\nu + \mu - 2i) \prod_{i=0}^{k} (\nu - \mu - 2i) \prod_{i=0}^{l+m} (n + \nu - i)} a_{0,0,0,0}.
\]

Now let \( M(-\nu) \) be the Verma module \( U(\mathfrak{g}) \otimes_{U(\mathfrak{p})} \mathbb{C}_{-\nu - \rho} \), where \( \mathbb{C}_{-\nu - \rho} \) is the \( \mathfrak{p} \)-module with \( \mathfrak{m} \) acting by \( d\xi_\mu \), \( \mathfrak{n} \) acting by 0 and \( a \) acting by \(-\nu - \rho\).
Let $u(\lambda, -\nu) = \sum_I a_I(\lambda, \nu) Y(I) \otimes 1$ be the $\chi$-Whittaker vector on the Verma module $M(-\nu)$, with $\chi$ as in Formula (2.4).

To obtain an explicit formula for the Whittaker vector $u(\lambda, -\nu)$ we change the parameter $\nu$ in Formulas (3.10) and (3.43) into $-\nu - \rho$, since $H$ acts by $-\nu(H) - \rho(H)$. Recall that for $G \simeq SU(n+1, 1)$ we have $\rho(H) = n+1$. With this parametrization and setting $a_{0,0,0} = 1$ in the case $n = 1$ and $a_{0,0,0,0} = 1$ in the case $n > 1$ we get

(3.44) 

$$a_{j,k,l}(\lambda, -\nu - \rho) = \frac{(-1)^{j+l}(\lambda/2)^{j+k+2l}}{j!k!l! \prod_{i=1}^{j+l}(\nu - \mu + n \frac{j+l+1}{2} + i) \prod_{i=1}^{k+l}(\nu + \mu + n \frac{j+l+1}{2} + i) \prod_{i=1}^{\nu}(\nu + i)},$$

(3.45)

$$a_{j,k,l,m}(\lambda, -\nu - \rho) = \frac{b_{j,k,l,m}(-\nu - \rho)(-1)^j \lambda^{j+k+2l+2m}}{j!k!(l+m)!l!(j+l+m) \prod_{i=1}^{j+l+m}(\nu - \mu + n \frac{j+l+1}{2} + i) \prod_{i=1}^{k+l+m}(\nu + \mu + n \frac{j+l+1}{2} + i) \prod_{i=1}^{(j+l+m)(\nu + i)}},$$

Furthermore, it will be convenient to multiply $u(\lambda, -\nu - \rho)$ by the normalizing factor $I(\nu) = \frac{1}{\Gamma(\nu + 1)} \Gamma(\frac{\nu + n}{2})^{-1} \Gamma(\frac{\nu + \mu + n}{2})^{-1} \Gamma(\nu + 1)^{-1}$ in order to obtain a holomorphic Whittaker vector $\tilde{u}(\lambda, -\nu - \rho)$. We may now state the main result in this section:

**Theorem 3.5.** Let $G$ be locally isomorphic with $SU(n+1, 1)$. Then a holomorphic $\chi$-Whittaker vector in $M(-\nu)$ is given by:

(1) If $n = 1$ then

(3.46) 

$$\tilde{u}(\lambda, -\nu - \rho) = \sum_{j,k,l \geq 0} \frac{(-1)^{j+l}(\lambda/2)^{j+k+2l}}{j!k!l! \Gamma\left(\frac{\nu - \mu}{2} + j + l + 1\right) \Gamma\left(\frac{\nu + \mu}{2} + k + l + 1\right) \Gamma(\nu + j + l + 1)} V_1^j V_2^k T_l,$$
Remark 1. We observe from the definition of $\tilde{u}(\lambda, -\nu - \rho) = \sum_{j,k,l,m \geq 0} \frac{(-1)^j b_{l,m}(-\nu - \rho) \lambda^{j+k+2l+2m} V_1^j V_2^k T^l q^m [\Gamma (\nu + l + m + 1)]^{-1}}{j! k! (l+m)! 4^l+2^m \Gamma \left( \frac{\nu - \mu + n}{2} + j + l + m + 1 \right) \Gamma \left( \frac{\nu + \mu + n}{2} + k + l + m + 1 \right)}$.

(3.47) $\tilde{u}(\lambda, -\nu - \rho) = \sum_{j,k,l,m \geq 0} \frac{(-1)^j b_{l,m}(-\nu - \rho) \lambda^{j+k+2l+2m} V_1^j V_2^k T^l q^m [\Gamma (\nu + l + m + 1)]^{-1}}{j! k! (l+m)! 4^l+2^m \Gamma \left( \frac{\nu - \mu + n}{2} + j + l + m + 1 \right) \Gamma \left( \frac{\nu + \mu + n}{2} + k + l + m + 1 \right)}$.

Remark 1. We observe from the definition of $\sigma(m, s, \nu)$ that

(3.48) $\sigma(m + s, s, -\nu - \rho) \pm \mu \left( \frac{m + 1 + s}{s + 1} \right)$

This together with (3.35) imply that $b_{l,m}(-\nu - \rho)$ is a polynomial in $\nu$ of degree $l$.

4. An explicit formula for the $\tau$-function.

Let $G$ be a Lie group locally isomorphic to $\text{SU}(n + 1, 1)$. We recall from §2 the definition of the $\tau$-function:

(4.1) $\tau(\chi_1, \chi, ua, \nu) = \sum_{I \in \mathbb{N}} a_I(-\nu) d\chi_1 (\text{Ad}(uas)^{-1} Y(I))$

where $u \in M$, $a \in A$ and $u(\lambda, -\nu) = \sum_I a_I(\lambda, -\nu) Y(I) \otimes 1$ is the $\chi$-Whittaker vector on the Verma module $M(-\nu)$. This function appears in the $\chi_1$-Fourier coefficient $D^\chi_{\chi_1}$ of the Poincaré series studied in [MW] and in the integral kernel of the Kuznetsov type formula in [MW1]. The case $\chi = \chi_1$ has special interest because the poles of the meromorphic continuation of $D^\chi_{\chi_1}(P, P, \nu)$ to $\mathbb{C}$ lie exactly at spectral parameters, that is, the nonzero eigenvalues of the Casimir operator $C$ on $L^2_0(\Gamma \backslash G/K)$ have the form $\nu_j(H)^2 - \rho(H)^2$, where $\nu_j$ is a pole of $\{D^\chi_{\chi_1}(P, P, \nu) \mid \chi \in (\Gamma \backslash N) - 1\}$ in the closed right half plane. Our main goal in this section will be to give an explicit formula for $\tau(\chi, \chi, ua, \nu)$.

For this purpose we need to compute $\text{Ad}(uas)^{-1} Y(I)$. If we take $s^* = \left( \begin{array}{ccc} 0 & 0 & i \\ 0 & I_n & 0 \end{array} \right)$, where $I_n$ denotes the $n \times n$ identity matrix, we have that $\text{Ad}(s^*)^{-1}Y_i = X'_i$, $\text{Ad}(s^*)^{-1}Y'_i = -X_i$, and therefore $\text{Ad}(s^*)^{-1}V_1 = 1/2(X'_1 - iX_1)$ and $\text{Ad}(s^*)^{-1}V_2 = 1/2(-X'_1 - iX_1)$. Also, since

$$\text{Ad}(u^{-1})V_1 = c_1 V_1 + d_1 V_2 + \sum_{j=2}^n (c_j Y_j + d_j Y'_j)$$

we obtain $\tau(\chi_1, \chi, ua, \nu)$.
for some coefficients $c_j$ and $d_j$ and using the fact that $\langle V_1, V_1 \rangle = \langle V_2, V_2 \rangle = 1/2$, we may write

$$\text{Ad}(u^{-1})V_1 = 2\langle V_1, \text{Ad}(u)V_1 \rangle V_1 + 2\langle V_1, \text{Ad}(u)V_2 \rangle V_2 + \sum_{j=2}^{n} (c_j Y_j + d_j Y')_j.$$  

In the same way we obtain

$$\text{Ad}(u^{-1})V_2 = 2\langle V_2, \text{Ad}(u)V_1 \rangle V_1 + 2\langle V_2, \text{Ad}(u)V_2 \rangle V_2 + \sum_{j=2}^{n} (c_j Y_j + d_j Y'_j).$$

As $d\chi(\text{Ad}(s^*)^{-1}V_1) = d\chi(\text{Ad}(s^*)^{-1}V_2) = -\frac{i\lambda}{2}$ we get the formula

$$(4.2) \quad d\chi(\text{Ad}(uas^*)^{-1}V_1^jV_2^k)$$

$$= a^{(j+k)\alpha} \left( -\frac{i\lambda}{2} \right)^{j+k} 2^{j+k} \langle V_1, \text{Ad}(u)(V_1 + V_2) \rangle^j \langle V_2, \text{Ad}(u)(V_1 + V_2) \rangle^k$$

$$= (-i\lambda a^\alpha)^{j+k} \langle V_1, \text{Ad}(u)(V_1 + V_2) \rangle^j \langle V_2, \text{Ad}(u)(V_1 + V_2) \rangle^k.$$

In the case of SU(2,1) the last formula can be simplified. If we take $u = u_\theta = \left( \begin{array}{cc} e^{i\theta} & 0 \\ 0 & e^{-2i\theta} \end{array} \right)$, then $\text{Ad}(u)V_1 = e^{3i\theta} V_1$, $\text{Ad}(u)V_2 = e^{-3i\theta} V_2$ and so

$$(4.3) \quad d\chi(\text{Ad}(uas^*)^{-1}V_1^jV_2^k)$$

$$= (-i\lambda a^\alpha)^{j+k} e^{3i(j-k)\theta} \langle V_1, V_1 \rangle^j \langle V_2, V_2 \rangle^k = \left( -\frac{i\lambda a^\alpha}{2} \right)^{j+k} e^{3i(j-k)\theta}.$$

Now, $T$ is $M$-invariant and it belongs to $g_{-2\alpha}$. Then $d\chi(\text{Ad}(s^*)^{-1}T^l) = 0$ if $l > 0$. Also, from the computations above we get:

$$d\chi(\text{Ad}(uas^*)^{-1}q^m) = a^{2m\alpha} d\chi(X_1^2)^m = (\lambda a^\alpha)^{2m}.$$

We shall use the following notation: Let be $\lambda \in i\mathbb{R}$, $a \in A$ and $u \in M$. We shall write

$$(4.4) \quad z = z(\lambda, a) = -i(\lambda/2)^2 a^\alpha,$$

$$(4.5) \quad \omega_i = \omega_i(u) = \langle V_i, \text{Ad}(u)(V_1 + V_2) \rangle, \quad i = 1, 2.$$

We note that $z \in i\mathbb{R}^0$, thus $\overline{z} = -z$. We also recall the definition of the generalized hypergeometric functions. We shall follow the notation of [SI]. Let $(a) = (a_1, a_2, \ldots, a_p)$ and $(b) = (b_1, \ldots, b_q)$. The generalized hypergeometric function $pFq((a); (b); y)$ is defined as follows:

$$pFq((a); (b); y) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{n!(b_1)_n \cdots (b_q)_n} y^n.$$
where \((c)_n = c(c+1) \ldots (c+n-1) = \frac{\Gamma(c+n)}{\Gamma(c)}\) if \(n \geq 1\) and \((c)_0 = 1\). We recall that if \(p \leq q\) then \(pF_q((a); (b); y)\) converges for every finite value of \(y\). For \(p = 0\) and \(q = 1\) then \(0F_1\) corresponds to the classical Bessel function \(I_\nu\), that is

\[0F_1 \left( ; b; z \right) = \Gamma(b)(z)^{b-1} I_{b-1} \left( 2z^\frac{1}{2} \right).\]

We now state the main result of this work.

**Theorem 4.1.** Let \(G\) be isomorphic to \(SU(n+1, 1)\), and let \(\tau(\chi, \chi, ua, \nu)\) be as in (4.1), \(u \in M, a \in A\) and \(\chi\) a character on \(N\) as in (2.4). Let \(z, \omega_1\) and \(\omega_2\) as be as in (4.4) and (4.5). If \(n = 1\) let \(u = u_\theta\) as in (4.3). Then

(4.6)

\[
\tau(\chi, \chi, u_\theta, \nu) = \sum_{j,k,m} \frac{(-1)^j \prod_{s=1}^{j+k} (\nu + s) z^{j+k} e^{3i(j-k)\theta}}{j! \prod_{s=1}^{j} (\nu + s) (\nu - \mu + s) \prod_{s=1}^{k} (\nu + s) (\nu + \mu + s) \prod_{s=1}^{m} (\nu + s)}
\]

If \(n > 1\) then

(4.7)

\[
\tau(\chi, \chi, ua, \nu) = \sum_{j,k,m \geq 0} \frac{(-1)^j (\pi/2)^{j+k} m! \omega_1^{j} \omega_2^{k}}{j!k!m! \prod_{s=1}^{j+k+m} \left( \frac{\nu - \mu + n - 1}{2} + s \right) \prod_{s=1}^{k+m} \left( \frac{\nu + \mu + n - 1}{2} + s \right) \prod_{s=1}^{m} (\nu + s)}
\]

(4.8)

\[
= \sum_{m \geq 0} \frac{(z/2)^{2m}}{m! (\nu + 1)_m \prod_{s=1}^{m} \left( \frac{\nu + \mu + n + 1}{2} \right) \prod_{s=1}^{m} \left( \frac{\nu - \mu + n + 1}{2} \right) m!} \cdot 0F_1 \left( \frac{\nu - \mu + 1}{2} + m + 1, \frac{\pi \omega_1}{2} \right) \cdot 0F_1 \left( \frac{\nu + \mu + 1}{2} + m + 1, \frac{2\omega_2}{2} \right)
\]
\[ (4.9) \]

\[
\frac{\Gamma \left( \frac{\nu - \mu + n + 1}{2} \right) \Gamma \left( \frac{\nu + \mu + n + 1}{2} \right)}{(2\pi)^2} \frac{(\pi \omega_1)^{n+1}}{(\pi \omega_2)^{n+1}} \cdot \sum_{m \geq 0} \frac{(\pi/2)^{2m} (\pi \omega_1)^{m} (\pi \omega_2)^{m}}{m!(\nu + 1)^m} I_{\nu - \mu - n - 1 + m} \left( (2\pi \omega_1)^{1/2} \right) I_{\nu + \mu - n - 1 + m} \left( (2\pi \omega_2)^{1/2} \right) .
\]

The first expressions for the \( \tau \)-functions are easily obtained from the definition of the \( \tau \)-function and the formulas in (4.2) and (4.3). Reordering these series and using the definition of the hypergeometric functions one gets the other alternative formulas.

Remark 2. The formula for the \( \tau \)-function in (4.6) suggests that if \( \mu \in \mathbb{N} \) is fixed, then there are a finite number of possible poles of the \( \tau \) function in \( \text{Re} \ \nu > 0 \). In particular, for \( SU(2,1) \) there should be simple poles at \( \nu = \mu - 2t, t \in \mathbb{N} \), provided that \( \mu - 2t > 0 \). We shall prove that these poles really exist by computing the residue at the point and showing that it is not zero.

Let \( \mu \in \mathbb{N} \), and let \( \nu \in \mathfrak{a}^* \), \( \nu > \rho \) such that \( \nu = \mu - 2t \) for some \( t \in \mathbb{N} \). Note that as \( \lambda \in i\mathbb{R} \) we have that \(-i(\lambda/2)^2a^\alpha = ix\), for some \( x \in \mathbb{R}^>0 \). Thus

\[
\text{Res}_{\nu = \nu_t} \tau(\chi, \chi, ua, \nu) = \lim_{\nu \to \nu_t} (\nu - \nu_t) \tau(\chi, \chi, ua, \nu)
\]

\[
= \lim_{\nu \to \nu_t} (\nu - \nu_t) \sum_{0 \leq j < t} \sum_{k \geq 0} \frac{(-1)^j j! k!}{(\nu - \mu + j + s) \prod_{s=1}^{k} \left( \frac{\nu + \mu}{2} + s \right)} \prod_{s=1}^{k} \left( \frac{\nu + s}{2} + s \right)
\]

\[
+ \lim_{\nu \to \nu_t} (\nu - \nu_t) \sum_{j \geq t} \sum_{k \geq 0} \frac{(-1)^j j! k!}{(\nu - \mu + j + s) \prod_{s=1}^{k} \left( \frac{\nu + \mu}{2} + s \right)} \prod_{s=1}^{k} \left( \frac{\nu + s}{2} + s \right)
\]
The first term in the last expression of the residue is zero because the infinite sum involved is holomorphic at \( \nu = \nu_t \). Then

\[
\text{Res}_{\nu=\nu_t} \tau(\chi, \chi, u, \nu) = \lim_{\nu \to \nu_t} \sum_{j \geq t, k \geq 0} \frac{2(-1)^j j! k! \prod_{s=1}^{t-1} (\nu - \mu + s) \prod_{s=t+1}^j (\nu - \mu + s) \prod_{s=1}^k (\nu + s) (\nu + \mu + s)}{j! k! \prod_{s=1}^{t-1} (\nu - \mu + s) \prod_{s=t+1}^j (\nu + s) (\nu + \mu + s)}.
\]

Putting \( \theta = \pi/6 \) and using that \((-1)^j = i^{-2j}\) we get

\[
(-1)^j (ix)^j k! e^{3i(j-k)\theta} = i^{-j+k} x^j k! e^{i\pi/2(j-k)} = x^{j+k}(-i)^{-j+k} = x^{j+k},
\]

which is a positive real number. Furthermore, the factor \( \prod_{s=1}^{t-1} (\nu - \mu + s) \) does not depend on \( j \) and \( k \). Then

(4.10) \[
\text{Res}_{\nu=\nu_t} \tau(\chi, \chi, u(\pi/6)a, \nu)
\]

\[
= \frac{2 (-1)^{t-1} (t-1)! \prod_{s=1}^{k} (\nu_t + s) x^{j+k}}{\prod_{s=1}^{t-1} (\nu_t + s) \prod_{s=t+1}^{j} (\nu_t + s) (\nu + \mu + s)}.
\]

Now, the series involved in the last expression is convergent and the terms are strictly greater than zero. This shows that the residue of the \( \tau \) function at \( \nu = \nu_t \) is not zero.

This proves the existence of simple poles of the \( \tau \) function for \( \text{SU}(2, 1) \) in \( \{ \text{Re} \nu > 0 \} \) at the real points \( \mu - 2, \mu - 4, \ldots \). The same should be true for \( \text{SU}(n + 1, 1) \), for \( n > 1 \) but we have not carried out the verification.

**References**


Received October 10, 2002.

FAAMF, Universidad Nacional de Cordoba
Ciudad Universitaria
Medina Allende s/n
5000 Cordoba
Argentina
E-mail address: kisbye@mate.uncor.edu
AN EXPLICIT ISOMORPHISM BETWEEN FLOER HOMOLOGY AND QUANTUM HOMOLOGY

GUANGCUN LU

We use Liu-Tian’s virtual moduli cycle methods to construct detailedly the explicit isomorphism between Floer homology and quantum homology for any closed symplectic manifold that was first outlined by Piunikhin, Salamon and Schwarz for the case of the semi-positive symplectic manifolds.

1. Introduction.

1.1. Background and motivation. It is one of the exciting mathematics achievements in the last few years that the Floer and quantum homologies were established for all closed symplectic manifolds (see [FuO, LiT, LiuT, R, Sie] and [HS2]). (Less general versions had been obtained before by Floer [F], Hofer-Salamon [HS1], Ruan-Tian [RT1] and McDuff-Salamon [McS].) The purpose of this paper is to construct detailedly an explicit ring isomorphism between them. Such an isomorphism was first outlined by Piunikhin, Salamon and Schwarz in the semi-positive case [PSSc]. Our argument is based on Liu-Tian’s virtual cycle methods in [LiuT1]-[LiuT3].

The isomorphism is necessary and convenient for studies of some symplectic topology problems, e.g., the topology and geometry of the group Ham(M,ω) of Hamiltonian automorphisms of a symplectic manifold (M,ω). Let G be the group of pairs (g, ̃g) consisting of a smooth loop g : S^1 → Ham(M,ω) such that g(0) = Id and a lift ̃g : ̃L(M) → ̃L(M) of the action of g to a covering of the space L(M) of contractible loops in M (see §1.2 below). In a beautiful paper [Se] by Seidel, for every pair (g, ̃g) ∈ ̃G there is assigned an automorphism HF∗(g, ̃g) of the Floer homology HF∗(M,ω); he constructed a homomorphism q from G to the group QH∗(M,ω) of homogeneous even-dimensional invertible elements of the quantum homology ring QH∗(M,ω) and proved his main result:

\[ HF∗(g, ̃g)(b) = \Psi^+(q(g, ̃g)) * PP b \]

for any (g, ̃g) ∈ ̃G and b ∈ HF∗(M,ω). Here *PP and Ψ+ are the ‘pair-of-pants’ product in HF∗(M,ω) and the canonical isomorphism QH∗(M,ω) ≅ HF∗(M,ω) constructed in [PSSc] respectively. A key step in the proof of
his main result is Theorem 8.2 on the page 1080 of [Se], whose proof was based on the arguments of [PSSc].

Schwarz [Sch3] defined and analyzed a bi-invariant metric on Ham($M, \omega$) with the construction of such an explicit isomorphism on a closed symplectic manifold $(M, \omega)$ with $c_1|_{\pi_2(M)} = \omega|_{\pi_2(M)} = 0$. Recently, Oh [Oh] obtained the corresponding results on arbitrary closed symplectic manifolds. As pointed out in §5.3 of [Oh] it would seem more natural to use the Piunikhin-Salamon-Schwarz map in the definition of his mini-max value function $\rho$. Entov [En] studied the relations between the K-area for Hamiltonian fibrations with a strongly semi-positive typical fiber $(M, \omega)$ over a surface with boundary and the Hofer geometry on the group Ham($M, \omega$). Such a ring isomorphism was used to obtain a key estimate in his work.

Since these applications used the Piunikhin-Salamon-Schwarz isomorphism in [PSSc] our detailed generalization to arbitrary closed symplectic manifolds may be used to generalize their results to the desired forms more directly and conveniently. Moreover the method to construct the ring isomorphism has actually more uses than the isomorphism itself because not only the isomorphism itself but also the map of Piunikhin-Salamon-Schwarz’s type in the chain level were used in some applications. The construction of another ring isomorphism was given by Liu-Tian [LiuT3] (a less general version was announced before by Ruan-Tian [RT2]). Without doubt different construction methods of the ring isomorphisms between Floer homology and quantum homology have respective advantages in the studies of different symplectic topology problems.

1.2. Outline and the main result. For a smooth nondegenerate time-dependent function $H : M \times \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ one may associate a family of the Hamiltonian vector fields $X_{H_t}$ by $\omega(X_{H_t}, \cdot) = -dH_t$ for $t \in \mathbb{R}$ and $H_t(\cdot) = H(t, \cdot)$. Let $\mathcal{P}(H)$ be the set of all contractible 1-periodic solutions of the Hamiltonian differential equation: $\dot{x}(t) = X_{H_t}(x(t))$. Denote by $\mathcal{J}(M, \omega)$ the space of all almost complex structures compatible with $\omega$. It determines a unique the first Chern class $c_1 = c_1(TM, J) \in H^2(M, \mathbb{Z})$ via any $J \in \mathcal{J}(M, \omega)$ ([Gr]). Let $\phi_{c_1}, \phi_\omega : H^S_2(M) \to \mathbb{R}$ be the homomorphisms by evaluations of $c_1$ and $\omega$ respectively. Here $H^S_2(M)$ denotes the image of $\pi_2(M)$ in $H_2(M; \mathbb{Z})$ under the Hurewicz homomorphism modulo torsion. As usual let $\mathcal{L}(M)$ be the set of all contractible loops $x \in C^\infty(S^1, M)$. Consider a pair $(x, v)$ consisting of $x \in \mathcal{L}(M)$ and a disk $v$ bounding $x$. Such two pairs $(x, v)$ and $(y, w)$ are called equivalent if $x = y$ and $\phi_{c_1}, \phi_\omega$ both take zero value on $v^\sharp(\cdot - w)$. Denote by $[x, v]$ the equivalence class of a pair $(x, v)$ and by $\mathcal{L}(M)$ the set of all such equivalence classes. Then the latter is a cover space of $\mathcal{L}(M)$ with the covering transformation group $\Gamma = H^S_2(M)/(\ker \phi_{c_1} \cap \ker \phi_\omega)$. Its action is given by $A \cdot [x, v] = [x, A^2_v v]$ for any $A \in \Gamma$, where $A^2_v v$ is understood as the connected sum of any representative of $A$ in $\pi_2(M)$ with $v$. In this paper we shall denote $[x, v]$ by $\bar{x}$ and $[x, v^\sharp A]$
by \( \bar{x}_A \) for \( A \in \Gamma \) if it is not necessary to point out the bounding disk \( v \) and no confusion occurs. Let \( \tilde{\mathcal{P}}(H) \) be the lifting of \( \mathcal{P}(H) \) in the space \( \tilde{\mathcal{L}}(M) \). It is exactly the critical set of the functional

\[
\mathcal{F}_H : \tilde{\mathcal{L}}(M) \to \mathbb{R}, \ [x,v] \mapsto -\int_D v^*\omega + \int_0^1 H(t,x(t))dt
\]
on \( \tilde{\mathcal{L}}(M) \). Given \( \bar{x}^\pm = [x^\pm, v^\pm] \in \tilde{\mathcal{P}}(H) \) denote by

\begin{equation}
\mathcal{M}(\bar{x}^-, \bar{x}^+; H, J)
\end{equation}

the space of all connecting trajectories \( u : \mathbb{R} \times S^1 \to M \) satisfying the equation \( \partial_s u + J'(u)\partial_t u + \nabla H(u) = 0 \) with boundary conditions \( \lim_{s \to \pm\infty} u(s,t) = x^\pm \) and \( [x^+, v^- u] = [x^+, v^+] \). After making a small generic perturbation of \( H_t \) outside some small neighborhood of the graph of the elements of \( \mathcal{P}(H) \) one may assume that for any two \( \bar{x}^\pm \in \tilde{\mathcal{P}}(H) \) the space \( \mathcal{M}(\bar{x}^-, \bar{x}^+; H, J) \) is either an empty set or a manifold of dimension \( \mu(\bar{x}^-) - \mu(\bar{x}^+) \) ([F]). Here \( \mu(\bar{x}) \) is the Conley-Zehnder index of \( \bar{x} \) ([SZ]).

Denote by \( \tilde{\mathcal{P}}_k(H) := \{ \bar{x} \in \tilde{\mathcal{P}}(H) \mid \mu(\bar{x}) = k \} \). Consider the chain complex whose \( k \)-th chain group \( C_k(H, J; \mathbb{Q}) \) consists of all formal sums \( \sum \xi_{\bar{x}} \cdot \bar{x} \) with \( \xi_{\bar{x}} \in \mathbb{Q} \) and \( \bar{x} \in \tilde{\mathcal{P}}_k(H) \) such that the set \( \{ \bar{x} \in \tilde{\mathcal{P}}_k(H) \mid \xi_{\bar{x}} \neq 0, \mathcal{F}_H(\bar{x}) > c \} \) is finite for any \( c \in \mathbb{R} \). Then \( C_*(H, J; \mathbb{Q}) = \bigoplus_k C_k(H, J; \mathbb{Q}) \) is a graded \( \mathbb{Q} \)-space of infinite dimension. However its dimension as a module over the Novikov ring \( \Lambda_{\omega} = \Lambda_{\omega}(\mathbb{Q}) \) is finite. Here \( \Lambda_{\omega}(\mathbb{Q}) \) is the collection of all formal sums \( \lambda = \sum \lambda_A \cdot e^A \) with \( \lambda_A \in \mathbb{Q} \) such that the set \( \{ A \in \Gamma \mid \lambda_A \neq 0, \omega(A) < c \} \) is finite for any \( c \in \mathbb{R} \). Its action on \( C_*(H, J; \mathbb{Q}) \) is defined by \( (\lambda \ast \xi)_{\bar{x}} = \sum_{A \in \Gamma} \lambda_A \xi(-A)_{\bar{x}} \) and the rank of \( C_*(H, J; \mathbb{Q}) \) over \( \Lambda_{\omega} \) is equal to \( \sharp \mathcal{P}(H) \). When \( (M, \omega) \) is a monotone symplectic manifold and \( \mu(\bar{x}^-) - \mu(\bar{x}^+) = 1 \) Floer proved that the manifold in (1.1) is compact and thus first established his Floer homology theory [F]. Later on his arguments were generalized to the semi-positive case by Hofer-Salamon [HS1] and Ono [O], and the case of the product of semi-positive symplectic manifolds by author [Lu1]. Note that the space in (1.1) is not compact in general case. It is this noncompactness that impedes the establishment of Floer homology theory on all closed symplectic manifolds. Let us outline Liu-Tian’s method to overcome this difficulty since we shall choose their method to realize our program. Replacing the space in (1.1) they considered \( \mathcal{M}(\bar{x}^-, \bar{x}^+; J, H) \) the space of the equivalence classes of all \( (J, H) \)-stable trajectories from \( \bar{x}^- \) to \( \bar{x}^+ \) (cf. Def. 2.1), and used it to construct a suitable relative virtual moduli cycle \( C(\mathcal{M}(\bar{x}^-, \bar{x}^+)) \) of dimension \( \mu(\bar{x}^-) - \mu(\bar{x}^+) - 1 \) (see §2.2). If \( \mu(\bar{x}^-) - \mu(\bar{x}^+) = 1 \) the virtual moduli cycle may determine a rational number \( \sharp(\mathcal{M}(\bar{x}^-, \bar{x}^+)) \) (see [LiuT1]). Then for each \( \xi = \sum_{\bar{x}} \xi_{\bar{x}} \bar{x} \) in
Given the quantum homology $C_k(H, J; \mathbb{Q})$ they defined

$$
\partial^F_k \xi = \sum_{\mu(\tilde{y})=k-1} \left[ \sum_{\mu(\tilde{x})=k} \sharp(C(\mathcal{M}^\nu(\tilde{x}, \tilde{y}))) \cdot \xi_{\tilde{x}} \right] \tilde{y}
$$

and proved it to be indeed a boundary operator. Let $HF_*(M, \omega; H, J, \nu; \mathbb{Q})$ be the homology of the above chain complex. Using Floer’s deformation ideas they proved that this homology is invariant under deformations and also isomorphic to $H_*(M, \mathbb{Q}) \otimes \Lambda_\omega$, i.e., the quantum homology of $M$.

For the construction of our isomorphism let us firstly fix a Morse function $h_0$ on $M$ and a small open neighborhood $\mathcal{O}(h_0)$ of it in $C^\infty(M)$ such that:

(i) For some $\epsilon > 0$ any two different points $a$ and $b$ of $\text{Crit}(h_0) = \{a_1, \ldots, a_m\}$ have a distance $d(a, b) > 4\epsilon$ with respect to some distance $d$ on $M$;

(ii) any two different critical points $a^h$ and $b^h$ of $h \in \mathcal{O}(h_0)$ have a distance $d(a^h, b^h) > 3\epsilon$;

(iii) each $h \in \mathcal{O}(h_0)$ has a unique critical point $a^h_i$ in each ball $B_d(a, \epsilon) = \{c \in M \mid d(a, c) < \epsilon\}$ and no other critical points (hence $\sharp(\text{Crit}(h)) = m$);

(iv) for any $h \in \mathcal{O}(h_0)$ the Morse index $\mu(a^h_i) = \mu(a_i)$, $i = 1, \ldots, m$;

(v) the function

$$
\mathcal{O}(h_0) \to B_d(a_1, \epsilon) \times \cdots \times B_d(a_m, \epsilon), \ h \mapsto (a^h_1, \ldots, a^h_m)
$$

is a smooth surjective map.

Take $h \in \mathcal{O}(h_0)$ and a Riemannian metric $g$ on $M$ such that $(h, g)$ is a Morse-Smale pair. As in [PSSc] we may use the solutions $\gamma : \mathbb{R} \to M$ of

$$
\dot{\gamma}(s) = -\nabla g h(\gamma(s))
$$

to construct a chain complex expression of the quantum homology $H_*(M, \mathbb{Q}) \otimes \Lambda_\omega$ as follows: For every integer $k$ let us denote by

$$
QC_k(M, \omega; h, g; \mathbb{Q})
$$

the set of all formal sums $\zeta = \sum_{\mu((a, A))=k} \zeta_{(a, A)} a \oplus A$ such that $\{(a, A) \in (\text{Crit}(h) \times \Gamma)_k \mid \zeta_{(a, A)} \neq 0, h(a) - \phi_\omega(A) > c\}$ is a finite set for all $c \in \mathbb{R}$. Here $(\text{Crit}(h) \times \Gamma)_k := \{(a, A) \in \text{Crit}(h) \times \Gamma \mid \mu((a, A)) := \mu(a) - 2c_1(A) = k\}$. The action of $\Lambda_\omega$ on $QC_*(M, \omega; h, g; \mathbb{Q})$ is given by

$$
\lambda \ast \zeta = \sum_{(b, B) \in \text{Crit}(h) \times \Gamma} \left( \sum_{2c_1(A) = \mu((b, B)) - k} \sum_{\lambda \oplus A \zeta((a, A) \cdot B, B, -A)} \lambda \ast A \right) \langle b, B \rangle.
$$
The boundary operator $\partial^Q_k : QC_k(M, \omega; h, g; \mathbb{Q}) \rightarrow QC_{k-1}(M, \omega; h, g; \mathbb{Q})$ is given by

$$\partial^Q_k(\langle a, A \rangle) = \sum_{\mu(b) = \mu(a)-1} n(a, b) \langle b, A \rangle,$$

where $n(a, b)$ is the oriented number of the solutions of (1.3) from $a$ to $b$. It is easily checked that $\lambda \ast \zeta \in QC_*(M, \omega; h, g; \mathbb{Q})$ and that the latter is a graded vector space over $\Lambda \omega (\mathbb{Q})$ according to the multiplication defined in (1.5), and that $\partial^Q$ is a boundary operator and also $\Lambda \omega (\mathbb{Q})$-linear with respect to the multiplication. Consequently, $(QC_*(M, \omega; h, g; \mathbb{Q}), \partial^Q)$ is a chain complex. Let us denote its homology by

$$QH_*(h, g; \mathbb{Q}) := H_*(QC_*(M, \omega; h, g; \mathbb{Q}), \partial^Q).$$

It is easy to derive from [Sch4] that there exists an explicit graded $\Lambda \omega$-module isomorphism between $QH_*(h, g; \mathbb{Q})$ and $H_*(M; \mathbb{Q}) \otimes \Lambda \omega (\mathbb{Q})$. To construct an explicit $\Lambda \omega$-module isomorphism between $QH_*(h, g; \mathbb{Q})$ and $HF_*(M, \omega; H, J, \nu; \mathbb{Q})$ we shall associated two rational numbers $n^+_{\nu}(a, \bar{x}##(-A))$ and $n^-_{\nu}(a, \bar{x}##(-A))$ in (3.2) to both $\langle a, A \rangle \in \text{Crit}(h) \times \Gamma$ and $\bar{x} \in \overline{P}(H)$, and then mimic [PSSc] to construct formally two maps on the levels of chains

$$\Phi(\langle a, A \rangle) = \sum_{\mu(\bar{x}) = \mu(a, A)} n^+_{\nu}(a, \bar{x}##(-A)) \bar{x},$$

$$\Psi(\bar{x}) = \sum_{\mu(\langle a, A \rangle) = \mu(\bar{x})} n^-_{\nu}(a, \bar{x}##(-A)) \langle a, A \rangle.$$ 

Indeed, in Remark 3.7 we shall show that $\Phi$ and $\Psi$ are $\Lambda \omega$-module chain homomorphisms from $QC_*(M, \omega; h, g; \mathbb{Q})$ to $C_*(H, J, \nu; \mathbb{Q})$ and from $C_*(H, J, \nu; \mathbb{Q})$ to $QC_*(M, \omega; h, g; \mathbb{Q})$ respectively. Our main result is:

**Theorem 1.1.** $\Phi$ induces a $\Lambda \omega$-module isomorphism

$$\Phi_* : QH_*(h, g; \mathbb{Q}) \rightarrow HF_*(M, \omega; H, J, \nu; \mathbb{Q})$$

with an inverse $\Psi_*$. 

**Remark 1.2.** As concluding remarks we point out that Theorems 3.1, 3.7 and 5.1 in [PSSc] may easily be extended to any closed symplectic manifold. Such an extension of Theorem 3.1 was actually carried out in [LiuT3]. Following the lines in [PSSc] and combing the methods in [LiuT3] with ones in this paper we easily complete the extension of Theorem 5.1 in [PSSc] to arbitrary closed symplectic manifolds (in fact, a long exercise). Hence the isomorphism in Theorem 1.1 is also a ring isomorphism.
2. The disk solution spaces and virtual moduli cycles.

2.1. Moduli space of stable disks. We begin with the disk solution spaces introduced in [PSSc]. For \( J \in \mathcal{J}(M, \omega) \) and \([x, v] \in \bar{\mathcal{P}}(H)\) let \( \mathcal{M}_+([x, v]; H, J) \) be the set of all smooth maps \( u : \mathbb{R} \times S^1 \to M \) such that
\[
\partial_s u(s, t) + J(u)(\partial_v u - \beta_{+}(s) X_H(t, u)) = 0,
\]
\[
u_{+}(v) : S^2 \to M \quad \text{represents a torsion homology class in} \quad H_2(M; \mathbb{Z}).
\]

Here a smooth cut-off function \( \beta_{+} : \mathbb{R} \to [0, 1] \) is given by
\[
\beta_{+}(s) = \begin{cases} 
0 & \text{as} \ s \leq 0, \\
1 & \text{as} \ s \geq 1.
\end{cases}
\]

Since \( E_+(u) < +\infty \), \( u \) extends over the end of \( s = -\infty \) by removable singularity theorem. So the connected union \( u_{+}(v) \) in (2.2) is well-defined. If \( \mathcal{M}_+([x, v]; H, J) \neq \emptyset \) its virtual dimension is \( \dim M - \mu([x, v]) \).

Correspondingly, for \([x, v] \in \bar{\mathcal{P}}(H)\) we denote by \( \mathcal{M}_-([x, v]; H, J) \) the set of all smooth maps \( u : \mathbb{R} \times S^1 \to M \) such that
\[
\partial_s u(s, t) + J(u)(\partial_v u - \beta_{-}(s) X_H(t, u)) = 0,
\]
\[
u_{-}(u) : S^2 \to M \quad \text{represents a torsion homology class in} \quad H_2(M; \mathbb{Z}).
\]

Similarly, if \( \mathcal{M}_-([x, v]; H, J) \neq \emptyset \) its virtual dimension is \( \mu([x, v]) \). As above we here have extended \( u \) over the end of \( s = +\infty \). Recall that:

**Definition 2.1 ([LiuT1]).** Let \((\Sigma, \bar{\ell})\) be a semistable \( \mathcal{F}\)-curve with \( z_- = z_1, \ldots, z_{N_p+1} = z_+ \), as those double points connecting the principal components (cf. Def. 3.1 in [LiuT2]). A continuous map \( f : \Sigma \setminus \{z_1, \ldots, z_{N_p+1}\} \to M \) is called a **stable \((J, H)\)-map** if there exist \( [x_i, v_i] \in \bar{\mathcal{P}}(H) \), \( i = 1, \ldots, N_p + 1 \), such that:

1. On each principal component \( P_i \) with cylindrical coordinate \((s, t)\) obtained by the identification \((P_i \setminus \{z_i, z_{i+1}\}; l_i) \equiv (\mathbb{R} \times S^1; \{t = 0\}), \)
\[
f^i_P = f|_{P_i \setminus \{z_i, z_{i+1}\}} \]
satisfies:

   (i) \( \partial_s f^i_P + J(f^i_P) \partial_v f^i_P + \nabla H(f^i_P) = 0 \),

   (ii) \( \lim_{s \to -\infty} f^i_P(s, t) = x_i(t) \) and \( \lim_{s \to +\infty} f^i_P(s, t) = x_{i+1}(t) \).
Definition 2.2. Given a \([x, v] \in \tilde{\mathcal{P}}(H)\) and a semistable \(\mathcal{F}\)-curve \((\Sigma, \mathcal{I})\), a continuous map \(f : \Sigma \setminus \{z_2, \ldots, z_{N_p+1}\} \to M\) is called a \textbf{stable} \((J, H)_+\)-\textbf{disk} with cap \([x, v]\) if there exist \([x_i, v_i] \in \tilde{\mathcal{P}}(H)\), \(i = 2, \ldots, N_p + 1\), with \([x, v] = [x_{N_p+1}, v_{N_p+1}]\), such that:

1. On each principal component \(P_i\) \((i > 1)\) with cylindrical coordinate \((s, t)\), \(f_i^P = f|_{P_i-(z_i, z_{i+1})}\) satisfies (i) (ii) in Definition 2.1(1), but \(f_i^P\) does (2.1) and \(f_i^P(+) = x_2\) and \(E_+(f_i^P) < +\infty\).
2. If \(N_p > 1\), for each \(i > 1\) as relative homology classes of \((M, x_{i+1})\), \([v_{i+1}] = [v_i] + [f_{i,j}^B] + \sum_j [f_{i,j}^B]\) and \([v_{2}] = [f_{i,j}^P] + \sum_j [f_{i,j}^B]\).
3. All requirements for the bubble components \(f_{i,j}^B\) in Definition 2.1(2)(4) still hold. Moreover, all homotopically trivial principal components \(f_i^P\) \((i > 1)\) are not free and do not appear in the next way.

Remark 2.3. Actually, \(f_i^P\) may not be constant. So, if \(f\) has at least two components then there exist at least two \textbf{nonconstant} components. This also holds for \(f_{N_p}^P\) in the following Definition 2.4. The energy of such a map is defined by

\[
(2.5) \quad E_+(f) = \sum_i \int\int_{\mathbb{R} \times S^1} |\partial_s f_i^P|^2 g_2 dsdt + \sum_{i,j} \int_{B_{i,j}} (f_{i,j}^B)^* \omega.
\]

Definition 2.4. Given a \([x, v] \in \tilde{\mathcal{P}}(H)\) and a \((\Sigma, \mathcal{I})\) as before, a continuous map \(f : \Sigma \setminus \{z_1, \ldots, z_{N_p}\} \to M\) is called a \textbf{stable} \((J, H)_-\)-\textbf{disk} with cap \([x, v]\) if there exist \([x_i, v_i] \in \tilde{\mathcal{P}}(H)\), \(i = 1, \ldots, N_p\), with \([x, v] = [x_1, v_1]\), such that:

1. On each principal component \(P_i\) \((i < N_p)\) with cylindrical coordinate \((s, t)\), \(f_i^P = f|_{P_i-(z_i, z_{i+1})}\) satisfies (i) (ii) in Definition 2.1(1), but \(f_{N_p}^P\) does (2.3) and \(f_{N_p}^P(-\infty) = x_{N_p}\) and \(E_-(f_{N_p}^P) < +\infty\).
2. If \(N_p > 1\), for each \(1 \leq i \leq N_p - 1\), as relative homology class of \((M, x_{i+1})\), \([v_i] = [v_{i+1}] + [f_{i,j}^P] + \sum_j [f_{i,j}^B]\) and \([v_{N_p}] = [f_{N_p}^P] + \sum_j [f_{N_p,j}^B]\).
(3) All assertions for the bubble components $f_{i,j}^B$ in Definition 2.1(2)(4) still hold. Moreover, all homotopically trivial principal components $f_i^P$ ($i < N_p$) are not free and do not appear in the next way.

We still define the energy $E_{\pm}(f)$ of such a map by the right side of (2.5). As before we may define their equivalence classes. Let us denote by $\langle f, \Sigma, l \rangle$ or simply $\langle f \rangle$ the equivalence class of $(f, \Sigma, l)$. The energy of $\langle f \rangle$ is defined by that of any representative of it. The direct computation shows that

$$E_{\pm}(f_{\pm}) \leq \mp \mathcal{F}_H([x, v]) + \max |H|. \quad (2.6)$$

The notions of the (effective) dual graph for the stable $(J, H)_{\pm}$-disks may also be defined with the same way as in [LiuT1]. Let $\mathcal{M}_{\pm}(\tilde{x}; H, J)$ be the spaces of the equivalence classes of all $(J, H)_{\pm}$-stable disks with cap $\tilde{x}$ respectively. By (2.6),

$$\pm \int_{D^2} v^* \omega \geq -2 \max |H| \quad \text{as} \quad \mathcal{M}_{\pm}(\tilde{x}; H, J) \neq \emptyset. \quad (2.7)$$

One may equip the weak $C^\infty$ topology on $\mathcal{M}_{\pm}(\tilde{x}; H, J)$ according to the definition given by (i)(ii)(iii) above Proposition 4.1 of [LiuT1] unless we allow the compact set $K$ in (ii) to be able to contain the double point $z_-$ (resp. $z_+$) on the chain of principal components of the domain of $\langle u_{\infty} \rangle \in \mathcal{M}_{\pm}(\tilde{x}; H, J)$ (resp. $\mathcal{M}_{-}(\tilde{x}; H, J)$). Carefully checking the proof of Proposition 4.1 in [LiuT1] we have:

**Proposition 2.5.** The spaces $\mathcal{M}_{\pm}(\tilde{x}; H, J) \supseteq \mathcal{M}_{\pm}(\tilde{x}; H, J)$ are compact and Hausdorff with respect to the weak $C^\infty$-topology.

Notice that we have two natural continuous maps

$$\text{EV}_{\pm}(\tilde{x}) : \mathcal{M}_{\pm}(\tilde{x}; H, J) \rightarrow M, \quad \langle f_{\pm} \rangle \mapsto f_{\pm}(z_{\pm}) \quad (2.8)$$

with respect to the weak $C^\infty$-topology. Correspondingly, the notions of the dual graphs of the stable $(J, H)_{\pm}$-disks with a cap $\tilde{x}$ may be introduced in similar ways. Let $D_{\pm}(\tilde{x})$ be the sets of their dual graphs respectively. Both are finite.

Now we are in the position to introduce the notions of positive and negative $L^p_k$-stable disks with cap $\tilde{x}$. As usual it is always assumed that $k - \frac{2}{p} \geq 1$. In principle, we may proceed as in [LiuT1]. For example, for each $D^+ \in D_{\pm}(\tilde{x})$ let $(f, \Sigma, l)$ be a stable $(J, H)_{\pm}$-disk as in Definition 2.2 and with the dual graph $D^+$. Then a positive $L^p_k$-stable disk with cap $\tilde{x}$ and of $D^+$ is a tuple $(\tilde{f}, \Sigma, l)$, where $\tilde{f} : \Sigma \setminus \{z_2, \ldots, z_{N_p+1}\} \rightarrow M$ is locally $L^p_k$ map such that (2) (3) in Definition 2.2 and the following are satisfied:

$$(1') \quad \tilde{f}^p = \tilde{f}|_{\tilde{f}^{-1}(z_i, z_{i+1})}(i > 1) \text{ satisfy (ii) in Definition 2.1(1) and suitable exponential decay condition along ends } z_i \text{ and } z_{i+1} \text{ as in [LiuT1], but}$$
\[ \mathcal{F}_1^P \] only satisfies \( \lim_{s \to +\infty} \mathcal{F}_1^P(s, t) = x_2(t) \) and the exponential decay condition along the end \( z_2 \).

We still define its energy \( E_+(f) \) by (2.5). The equivalence class of it can also be defined similarly. Denote by \( \mathcal{B}^p_{\pm}(\widetilde{x}; H) \) the sets of equivalence classes of those \( L^p_k \)-stable disks with cap \( \widetilde{x} \) and of the dual graphs in \( D_{\pm}(\widetilde{x}) \), and whose energy are less than \( +\mathcal{F}_H(\widetilde{x}) + \max \{H\} + 1 \) (because of (2.6)). As in [LiuT1] one can equip the strong \( L^p_k \)-topology on small neighborhoods \( \mathcal{W}_{\pm}(\widetilde{x}; H, J) \) of \( \mathcal{M}_{\pm}(\widetilde{x}; H, J) \) in \( \mathcal{B}^p_{\pm}(\widetilde{x}; H) \) and prove that it is equivalent to the above weak \( C^\infty \)-topology on \( \mathcal{M}_{\pm}(\widetilde{x}; H, J) \). Once these are well-defined we can use Liu-Tian’s method to construct the virtual moduli cycles

\[
\mathcal{M}^+_{\nu}(\widetilde{x}; H, J) = \sum_{I \in \mathcal{N}(\widetilde{x})_+} \frac{1}{|I|} \left\{ \pi^+_I : \mathcal{M}^+_{\nu}^{\nu}(\widetilde{x}; H, J) \to \mathcal{W}_+(\widetilde{x}; H, J) \right\}
\]

of dimension \( \dim M - \mu(\widetilde{x}) \) in \( \mathcal{W}_+(\widetilde{x}; H, J) \), and

\[
\mathcal{M}^-_{\nu}(\widetilde{x}; H, J) = \sum_{I \in \mathcal{N}(\widetilde{x})_-} \frac{1}{|I|} \left\{ \pi^-_I : \mathcal{M}^-_{\nu}^{\nu}(\widetilde{x}; H, J) \to \mathcal{W}_-(\widetilde{x}; H, J) \right\}
\]

of dimension \( \mu(\widetilde{x}) \) in \( \mathcal{W}_-(\widetilde{x}; H, J) \) (cf. [LiuT1, LiuT2] and [LiuT3]). It should also be pointed out that the maps \( \text{EV}_{\nu} \) in (2.8) can be naturally extended onto the spaces \( \mathcal{W}_{\pm}(\widetilde{x}; H, J) \). By composing them with the obvious finite-to-one maps from \( \mathcal{M}^\nu_{\nu}(\widetilde{x}; H, J) \) to \( \mathcal{W}_{\pm}(\widetilde{x}; H, J) \) that forget the parameterization we get two continuous and stratawise smooth evaluations

\[
\text{EV}^\nu_{\nu}(\widetilde{x}) : \mathcal{M}^\nu_{\nu}(\widetilde{x}; H, J) \to M.
\]

Notice that the choices of different small \( \nu \) give the cobordant virtual moduli cycles \( \mathcal{M}^\nu_{\nu}(\widetilde{x}; H, J) \) and corresponding evaluations \( \text{EV}^\nu_{\nu} \).

However, notice that the boundary operator \( \partial^F \) in (1.2) depends on the choice of \( \nu \). As pointed out below Remark 4.2 of [LiuT1], for \( \partial^F \) in (1.2) being indeed a boundary operator the choices of \( \nu \) in all relative virtual moduli cycles \( \mathcal{M}^\nu(\widetilde{x}, \widetilde{y}) \) (\( \mu(\widetilde{x}) - \mu(\widetilde{y}) \leq 2 \)) must satisfy suitable compatible conditions, i.e.,

\[
\partial \left( C \left( \mathcal{M}^\nu(\widetilde{x}, \widetilde{y}) \right) \right) = \sum_{\mu(\widetilde{z}) = \mu(\widetilde{x}) - 1} C \left( \mathcal{M}^\nu(\widetilde{x}, \widetilde{z}) \right) \times C \left( \mathcal{M}^\nu(\widetilde{y}, \widetilde{z}) \right)
\]

holds for all pairs \((\widetilde{x}, \widetilde{y})\) with \( \mu(\widetilde{y}) - \mu(\widetilde{x}) = 2 \). Now in order to guarantee that the maps \( \Phi \) and \( \Psi \) constructed in (1.7) and (1.8) commute with \( \partial^F \) and the boundary operator \( \partial^Q \) in (1.6), we also need to carefully choose \( \nu \) in (2.9) and (2.10) so that they are compatible with all \( \nu \) chosen in the definition of \( \partial^F \) in (1.2).
2.2. Compatible virtual moduli cycles. In this subsection we shall first complete Liu-Tian’s arguments in detail, i.e., proving (2.12), and then outline how to construct all virtual moduli cycles $\overline{\mathcal{M}}^\nu_\pm (\bar{x}; H, J)$ compatible with all relative virtual moduli cycles in (1.2). For convenience of the later proof we need to recall briefly the construction of the relative virtual cycle $C (\overline{\mathcal{M}}^\nu (\bar{x}, \bar{y}))$ in [LiuT1].

**Step 1**: Local construction. For a representative $f$ of $\langle f \rangle \in \overline{\mathcal{M}}(\bar{x}, \bar{y}; H, J)$ one may construct a stratified Banach orbifold chart $(\tilde{W}(f), \Gamma_f, \pi_f)$ around $\langle f \rangle$ in $\mathcal{B}(\bar{x}, \bar{y}) = \mathcal{B}^\nu_k(\bar{x}, \bar{y})$, where $\Gamma_f = \text{Aut}(f)$. There exists a natural stratified Banach bundle $\tilde{L}(f) \to \tilde{W}(f)$ with a stratawise smooth right $\Gamma_f$-action such that the usual $\partial J, H$-operator gives rise to a $\Gamma_f$-equivariant stratawise smooth section of this bundle, still denoted by $\tilde{\partial}$ $\mathcal{J}, H$. Let $I$ denote the dual graph of $f$ and $R(f) \subset (\tilde{L}(f))_f$ the cokernel $\text{coker} (\tilde{\partial} J, H(f))$. Take a smooth cut-off function $\beta_f$ supported outside of the $\epsilon$-neighborhood of double points of the domain $\Sigma_f$. Then each $\nu \in R_\epsilon(f) := \{ \beta_f \cdot \xi | \xi \in R(f) \}$ may naturally determine a section of the bundle $\tilde{L}(f) \to \tilde{W}(f)$, denoted by $\tilde{\nu}$, such that for each $g \in \tilde{W}(f)$ the support of $\tilde{\nu}(g)$ in $(\tilde{L}(f))(g)$ is away from the gluing region of the domain $\Sigma_g$ of $g$.

**Step 2**: Global construction. By the compactness of $\overline{\mathcal{M}}(\bar{x}, \bar{y}; H, J)$ one can choose finite points $\langle f_1 \rangle, \ldots, \langle f_m \rangle$ such that the union $W := \bigcup_{i=1}^m W(\langle f_i \rangle)$ is an open neighborhood of $\overline{\mathcal{M}}(\bar{x}, \bar{y}; H, J)$ and $\mathcal{L} := \bigcup_{i=1}^m \mathcal{L}(\langle f_i \rangle)$ is an orbifold bundle over $W$. Let $\mathcal{N}(\bar{x}, \bar{y})$ be the set of all subsets $I = \{i_1, \ldots, i_l\}$ of $\{1, \ldots, m\}$ with $W_I := \cap_{i \in I} W(\langle f_i \rangle) \neq \emptyset$. Let $\pi_I : \tilde{W}(f_i) \to W(\langle f_i \rangle)$ be the natural projections. For each $I \in \mathcal{N}(\bar{x}, \bar{y})$ they defined the group $\Gamma_I := \prod_{i \in I} \Gamma_{f_i}$ and the fiber product

\begin{equation}
\tilde{W}_I := \left\{ (u_i)_{i \in I} \in \prod_{i \in I} \tilde{W}(f_i) \mid \pi_I(u_i) = \pi_I(u_j) \forall i, j \in I \right\}.
\end{equation}

Then the projection $\pi_I : \tilde{W}_I \to W_I$ has covering group $\Gamma_I$. Moreover, for $J \subset I \in \mathcal{N}(\bar{x}, \bar{y})$ there is an obvious projection $\pi_I^J : \tilde{W}_I \to \tilde{W}_J$ satisfying the relation $\pi_J \circ \pi_I^J = \pi_I$. Repeating the same construction from $\tilde{L}(f_i)$ one obtains the bundles $\tilde{L}_I$ and thus a system of bundles $(\tilde{L}_I, \tilde{W}_I) = \left\{ (\tilde{L}_I^I, \tilde{W}_I^I), \pi_I^J, \pi_J \mid J \subset I \in \mathcal{N}(\bar{x}, \bar{y}) \right\}$. Take open sets $W_1^I \subset W(\langle f_i \rangle)$, $i = 1, \ldots, m$ and the pairs of open sets $W_1^I \subset U_1^I$, $i,j = 1, \ldots, m$, such that $\bigcup_{i=1}^m W_1^I$ still contains $\overline{\mathcal{M}}(\bar{x}, \bar{y}; J, H)$ and that

\begin{equation}
W_1^I \subset U_1^I \subset W_2^I \subset \cdots \subset W_m^I \subset W(\langle f_i \rangle).
\end{equation}
In [LiuT1], for each $I \in \mathcal{N}(\bar{x}, \bar{y})$ with cardinal number $|I| = k$, they defined

\begin{equation}
V_I = \cap_{i \in I} W_i^k \cup \cup_{|J| > k} \text{Cl}(\cap_{j \in J} U_j^k)
\end{equation}

and proved that the open covering $\{V_I | I \in \mathcal{N}(\bar{x}, \bar{y})\}$ of $\mathcal{M}(\bar{x}, \bar{y}; H, J)$ satisfies

\begin{equation}
V_I \subset W_I \forall I \in \mathcal{N}(\bar{x}, \bar{y}), \quad \text{and} \quad \text{Cl}(V_I) \cap \text{Cl}(V_J) \neq \emptyset \quad \text{only if} \quad I \subset J \quad \text{or} \quad J \subset I.
\end{equation}

Set $\tilde{V}_I = (\pi_I)^{-1}(V_I)$ and $\tilde{E}_I = (\pi_I)^{-1}(\mathcal{L}|_{V_I})$, one gets a system of bundles

\begin{equation}
(\tilde{E}^\Gamma, \tilde{V}^\Gamma) = \left\{ \left( \tilde{E}_I^{\Gamma}, \tilde{V}_I^{\Gamma} \right), \pi_I, \Gamma, \pi_J^I | J \subset I \in \mathcal{N}(\bar{x}, \bar{y}) \right\}.
\end{equation}

Taking $\Gamma_{f_i}$-invariant stratwise smooth cut-off function $\gamma(f_i)$ on $\tilde{W}(f_i)$ such that

\begin{equation}
\gamma(f_i) = 1 \quad \text{in} \quad \pi_i^{-1}(W_i^m),
\end{equation}

then each $\nu_i \in R_\epsilon(f_i)$ determines a smooth global section $\bar{\nu}_i$ of $(\tilde{E}^\Gamma, \tilde{V}^\Gamma)$. Set

\begin{equation}
R^\epsilon_\delta(\{f_i\}) = \{ \nu \in \oplus_{i=1}^m R_\epsilon(f_i) | |\nu| < \delta \}
\end{equation}

for a small $\delta > 0$. The bundle system $(\tilde{E}^\Gamma \times R^\epsilon_\delta(\{f_i\}), \tilde{V}^\Gamma \times R^\epsilon_\delta(\{f_i\}))$ has a well-defined global section

\begin{equation}
\bar{\partial}_{I,J} + \epsilon : (u_I, \nu) \mapsto \bar{\partial}_{I,J} u_I + \sum_{i=1}^m (\bar{\nu}_i)_{I}(u_I)
\end{equation}

for any $(u_I, \nu) = (u_I, (\nu_1, \ldots, \nu_m)) \in \tilde{V}_I \times R^\epsilon_\delta(\{f_i\})$. Moreover, each $\nu \in R^\epsilon_\delta(\{f_i\})$ yields a smooth section $\bar{\partial}^\epsilon_{I,J} = \{(\bar{\partial}^\epsilon_{I,J})_I | I \in \mathcal{N}(\bar{x}, \bar{y})\}$ of $(\tilde{E}^\Gamma, \tilde{V}^\Gamma)$,

\begin{equation}
(\bar{\partial}^\epsilon_{I,J})_I(u_I) = \bar{\partial}_{I,J} u_I + \sum_{i=1}^m (\bar{\nu}_i)_{I}(u_I) \quad \forall u_I \in \tilde{V}_I.
\end{equation}

**Theorem 2.6 ([LiuT1])**. The section in (2.20) is smooth and transversal to the zero section. Therefore when $\delta > 0$ is small enough, for a generic choice of $\nu \in R^\epsilon_\delta(\{f_i\})$ the section $\bar{\partial}^\epsilon_{I,J}$ is transversal to the zero section. Thus the family of perturbed moduli spaces $\tilde{\mathcal{M}}^\epsilon = \tilde{\mathcal{M}}^\epsilon(\bar{x}, \bar{y}) = \{ \tilde{\mathcal{M}}^\epsilon = (\bar{\partial}^\epsilon_{I,J})^{-1}(0) | I \in \mathcal{N}(\bar{x}, \bar{y}) \}$ is compatible in the sense that $\pi_I^1(\tilde{\mathcal{M}}^\epsilon) = \tilde{\mathcal{M}}^\epsilon \cap (\text{Im} \pi_I^1)$ for all $J \subset I \in \mathcal{N}(\bar{x}, \bar{y})$.

Let $W_i^m$ be given by (2.14). For sufficiently small $\nu$ we can also require that

\begin{equation}
\tilde{\mathcal{M}}^\epsilon \subset \cap_{i \in I} \pi_i^{-1}(W_i^m).
\end{equation}
Let $\tilde{\mathcal{M}}_I^{\mu,DT}$ (resp. $\tilde{\mathcal{M}}_I^{\mu,DB}$) be the top strata (resp. the strata of "broken" connecting orbits) of $\mathcal{M}_I^\mu$. Then $\tilde{\mathcal{M}}_I^{\mu,c} := \tilde{\mathcal{M}}_I^{\mu,DT} \cup \tilde{\mathcal{M}}_I^{\mu,DB}$ is a smooth manifold of dimension $\mu(x) - \mu(y) - 2$ and with boundary $\partial \tilde{\mathcal{M}}_I^{\mu,c} = \tilde{\mathcal{M}}_I^{\mu,DB}$. Formally writing

$$C(\mathcal{M}^\nu(x,y)) = \sum_{I \in \mathcal{N}(x,y)} \frac{1}{|I|} \tilde{\mathcal{M}}_I^{\nu,c}, \quad \mathcal{M}^\nu(x,y) = \sum_{I \in \mathcal{N}(x,y)} \frac{1}{|I|} \tilde{\mathcal{M}}_I^\nu,$$

the former was called the relative moduli cycle in [LiuT1]. Later we call the compatible family $\tilde{\mathcal{M}}^\nu(x,y)$ in Theorem 2.6 a derived family for $C(\mathcal{M}^\nu(x,y))$ and $\tilde{\mathcal{M}}^\nu(x,y)$. It is clear that $C(\mathcal{M}^\nu(x,y)) = \tilde{\mathcal{M}}^\nu(x,y)$ in the case $\mu(x) - \mu(y) \leq 2$. For the sake of clearness $C(\mathcal{M}^\nu(x,y))$ will be denoted by $C\left(\mathcal{M}^{\nu(\tilde{y})}(x,y)\right)$ below. If $\mu(x) - \mu(y) = 2$, then for every $\tilde{z} \in \tilde{\mathcal{P}}(H)$ with $\mu(x) - \mu(\tilde{z}) = 1$ one has also the associated relative moduli cycles $C\left(\mathcal{M}^{\nu(\tilde{y})}(\tilde{z},y)\right)$ and $C\left(\mathcal{M}^{\nu(\tilde{y})}(\tilde{z},y)\right)$. The relative virtual moduli cycles satisfying (2.12) are called compatible.

Proof of (2.12). To construct such relative virtual cycles, note that there are only finitely many $\tilde{z} \in \tilde{\mathcal{P}}(H)$, saying $\tilde{z}_1, \ldots, \tilde{z}_r$, such that $\mu(\tilde{z}_q) = \mu(x) - 1$ and $\mathcal{M}(\tilde{z}_q, H, J) \neq \emptyset$. Let $(f_s^{(1q)}) \in \tilde{\mathcal{M}}(\tilde{x}, \tilde{z}_q, H, J)$, $(f_t^{(2q)}) \in \tilde{\mathcal{M}}(\tilde{z}_q, \tilde{y}, H, J)$, $s = 1, \ldots, m_q$, $t = 1, \ldots, n_q$, be finite points from which one can construct the relative virtual moduli cycles

\begin{align}
C(\tilde{\mathcal{M}}^{\nu(\tilde{z}_q)}(\tilde{x}, \tilde{z}_q)) &= \sum_{I \in \mathcal{N}(\tilde{x}, \tilde{z}_q)} \frac{1}{|I(1q)|} \tilde{\mathcal{M}}^{\nu(\tilde{z}_q),c}, \\
C(\tilde{\mathcal{M}}^{\nu(\tilde{z}_q)}(\tilde{z}_q, \tilde{y})) &= \sum_{I \in \mathcal{N}(\tilde{z}_q, \tilde{y})} \frac{1}{|I(2q)|} \tilde{\mathcal{M}}^{\nu(\tilde{z}_q),c}.
\end{align}

For the future convenience we assume that for $q = 1, \ldots, r$,

\begin{align}
W^1((f_s^{(1q)})) &\subset U^1((f_s^{(1q)})) \subset U^1((f_s^{(1q)})) \\
&\subset W^m((f_s^{(1q)})) \subset W((f_s^{(1q)})), \quad s = 1, \ldots, m_q,
\end{align}

and

\begin{align}
W^1((f_t^{(2q)})) &\subset U^1((f_t^{(2q)})) \subset U^1((f_t^{(2q)})) \\
&\subset W^m((f_t^{(2q)})) \subset W((f_t^{(2q)})), \quad t = 1, \ldots, n_q,
\end{align}

are respectively the open sets as defined in (2.14) that are used to construct the relative virtual moduli cycles in (2.23) and (2.24) above. By (2.15)-(2.17) we get the corresponding bundle systems $(\tilde{E}^{(1q)}, \tilde{V}^{(1q)})$ and $(\tilde{E}^{(2q)}, \tilde{V}^{(2q)})$,
$q = 1, \ldots, r$. As in Theorem 2.6 let
\begin{align}
(2.27) \quad \nu(xz_q) &= \bigoplus_{s=1}^{m_q} \nu(xz_q)_s \in R_t^s(\{f_s^{(1q)}\}) \\
\nu(z_qy) &= \bigoplus_{t=1}^{n_q} \nu(z_qy)_t \in R_t^s(\{f_t^{(2q)}\}) 
\end{align}
be such that the smooth section $\tilde{\partial}_j^{(\varepsilon z_q)}$ of the bundle system $(\tilde{E}^1, \tilde{V}_1^{(1q)})$ and that $\tilde{\partial}_j^{(\varepsilon qy)}$ of $(\tilde{E}^2, \tilde{V}_2^{(2q)})$ are transversal to the zero section respectively. Here $\nu(xz_q)_s \in R_s(f_s^{(1q)})$ and $\nu(z_qy)_t \in R_t(f_t^{(2q)})$, $s = 1, \ldots, m_q$ and $t = 1, \ldots, n_q$. In fact, as in (2.21) we can also require $\nu(xz_q)$ and $\nu(z_qy)$ so small that
\begin{align}
\tilde{M}_I^{(\varepsilon z_q)} &\subset \cap_{s \in I}(x^{(1q)})^{-1}(W_{m_q}(f_s^{(1q)})) \forall I \in \mathcal{N}(xz_q), \\
\tilde{M}_I^{(\varepsilon qy)} &\subset \cap_{t \in I}(x^{(2q)})^{-1}(W_{n_q}(f_t^{(2q)})) \forall I \in \mathcal{N}(z_qy).
\end{align}
For $q = 1, \ldots, r$ we set $f_{st}^{(q)} := f_s^{(1q)} \cup f_t^{(2q)}$, $s = 1, \ldots, m_q$, $t = 1, \ldots, n_q$. Clearly, these $(f_{st}^{(q)})$ belong to $\tilde{M}(x; y; J, H)$. Moreover, the automorphism group $\Gamma_{st}^{(q)}$ of $f_{st}^{(q)}$ can be identified with the product $\Gamma_s^{(1q)} \times \Gamma_t^{(2q)}$ of the automorphism group $\Gamma_s^{(1q)}$ of $f_s^{(1q)}$ and $\Gamma_t^{(2q)}$ of $f_t^{(2q)}$. By the construction of the local uniformizer in §2 of [Liu1] we easily construct a uniformizer $\tilde{W}(f_{st}^{(q)})$ such that
\begin{align}
(2.28) \quad \tilde{W}(f_{st}^{(1q)}) \cup \tilde{W}(f_{st}^{(2q)}) \subset \tilde{W}(f_{st}^{(q)})
\end{align}
and that the restriction of $\Gamma_{st}^{(q)}$-action over $\tilde{W}(f_{st}^{(q)})$ to $\tilde{W}(f_{st}^{(1q)}) \cup \tilde{W}(f_{st}^{(2q)})$ is exactly that of $\Gamma_s^{(1q)} \times \Gamma_t^{(2q)}$ over $\tilde{W}(f_{st}^{(1q)}) \cup \tilde{W}(f_{st}^{(2q)})$ in the obvious way, where $\tilde{W}(f_{st}^{(1q)}) \cup \tilde{W}(f_{st}^{(2q)})$ denotes the set of all join functions at $z_q$ of functions in $\tilde{W}(f_{st}^{(1q)})$ and $\tilde{W}(f_{st}^{(2q)})$. (One may increase $m_q, n_q$ and shrink $\tilde{W}(f_{st}^{(1q)})$ and $\tilde{W}(f_{st}^{(2q)})$ if necessary.) Since each $\mathcal{M}(x; z_q; J, H) \cup \mathcal{M}(x; y; J, H)$ is compact and $\cup_{q=1}^{r} \mathcal{M}(x; z_q; J, H) \cup \mathcal{M}(x; y; J, H)$ are disjoint unions we can require that the above uniformizers $\tilde{W}(f_{st}^{(q)})$ satisfy
\begin{align}
(2.29) \quad \tilde{W}(f_{st}^{(q)}) \cap \tilde{W}(f_{st'}^{(q)}) = \emptyset \quad \forall q \neq q'.
\end{align}
By the definition of the index set $\mathcal{N}(x, y)$ above (2.13) it easily follows from (2.29) that the corresponding index set $\mathcal{N}^0(x, y)$ with the collection $\{\tilde{W}(f_{st}^{(q)}) \mid 1 \leq s \leq m_q, 1 \leq t \leq n_q, 1 \leq q \leq r\}$ must have the following form:
\begin{align}
(2.30) \quad \mathcal{N}^0(x, y) = \cup_{q=1}^{r} \mathcal{N}(x, z_q) \times \mathcal{N}(z_q, y).
\end{align}
Notice that every $W((f_{st}^{(q)}))$ determines an open neighborhood $W((f_{st}^{(q)}))$ of $(f_{st}^{(q)})$ in $\mathcal{B}(x, z_q)$ by
\begin{align}
\left\{ (g_1) \in \mathcal{B}(x, z_q) \mid \exists (g_2) \in \mathcal{B}(z_q, y) \text{ s.t. } (g_1, g_2) \in \mathcal{B}(x, y) \cap W(f_{st}^{(q)}) \right\}
\end{align}
and that $W(\langle f^{(q)}_{st} \rangle)_{2}$ of $\langle f^{(2q)}_{t} \rangle$ in $\mathcal{B}(\tilde{z}, \tilde{y})$ by

$$\{ (g_{2}) \in \mathcal{B}(\tilde{z}, \tilde{x}) \mid \exists (g_{1}) \in \mathcal{B}(\tilde{x}, \tilde{z}) \text{ s.t. } (g_{1} \circ g_{2}) \in \mathcal{B}(\tilde{x}, \tilde{y}) \cap W(f^{(q)}_{st}) \}.$$  

Thus (2.28) implies that

$$W(\langle f^{(1q)}_{s} \rangle) \subset W(\langle f^{(q)}_{st} \rangle)_{1} \text{ and } W(\langle f^{(2q)}_{t} \rangle) \subset W(\langle f^{(q)}_{st} \rangle)_{2}$$

for all $q, s, t$. By this, (2.25) and (2.26) we can, as in (2.14), choose pairs of open sets $W^{j}(\langle f^{(q)}_{st} \rangle) \subset U^{j}(\langle f^{(q)}_{st} \rangle), j = 1, \ldots, m = \sum_{q=1}^{r} m_{q} n_{q}$, such that

$$W^{1}(\langle f^{(q)}_{st} \rangle) \subset U^{1}(\langle f^{(q)}_{st} \rangle) \subset W^{2}(\langle f^{(q)}_{st} \rangle) \ldots \subset W^{m}(\langle f^{(q)}_{st} \rangle) \subset W(\langle f^{(q)}_{st} \rangle)$$

and that $W^{m_{q}}(\langle f^{(1q)}_{s} \rangle) \subset W^{1}(\langle f^{(q)}_{st} \rangle)_{1}$ and $W^{m_{q}}(\langle f^{(2q)}_{t} \rangle) \subset W^{1}(\langle f^{(q)}_{st} \rangle)_{2}$ for all $q, s, t$. As in (2.15) and (2.17) we use (2.30) and (2.31) to construct a bundle system

$$(\tilde{E}, \tilde{V}) = \left\{ (\tilde{E}_{I_{1}\times I_{2}}, \tilde{V}_{I_{1}\times I_{2}}) \mid I_{1} \times I_{2} \in \mathcal{N}^{0}(\tilde{x}, \tilde{y}) \right\}.$$  

As in (2.18) we take $\Gamma^{(q)}_{st}$-invariant smooth cut-off function $\gamma(\langle f^{(q)}_{st} \rangle)$ on $\tilde{W}(\langle f^{(q)}_{st} \rangle)$ such that $\gamma(\langle f^{(q)}_{st} \rangle) = 1$ in $(\pi^{q}_{st})^{-1}(W^{m}(\langle f^{(q)}_{st} \rangle))$. Note that the supports of $\nu(\tilde{x}z_{q})$ and $\nu(\tilde{z}y_{t})$ in (2.27) are away from double points on their domains. So the support of $\nu(\tilde{z}q)_{st} := \nu(\tilde{x}z_{q}) \nu(\tilde{z}q)_{t}$ is away from all double points of the domain of $\langle f^{(q)}_{st} \rangle$ and thus $\nu(\tilde{z}q)_{st}$ sits in the fibre of $\tilde{E}(\langle f^{(q)}_{st} \rangle)$ at $f^{(q)}_{st}, L_{k-1}(\lambda^{0,1}(\langle f^{(q)}_{st} \rangle^{*}TM))$. As before we can use $\gamma(\langle f^{(q)}_{st} \rangle)$ and $\nu(\tilde{z}q)_{st}$ to get a global section

$$\nu(\tilde{z}q)_{st}^{q} = \left\{ (\nu(\tilde{z}q)_{st})_{I_{1}\times I_{2}} \mid I_{1} \times I_{2} \in \mathcal{N}^{0}(\tilde{x}, \tilde{y}) \right\}$$

of the bundle system $(\tilde{E}, \tilde{V})$. Let us set

$$\nu(\tilde{x}y)^{q} := \sum_{q=1}^{r} \sum_{s=1}^{m_{q}} \sum_{t=1}^{n_{q}} \nu(\tilde{x}y)^{q}_{st} \text{ and } \nu(\tilde{x}y)^{q} := \sum_{q=1}^{r} \sum_{s=1}^{m_{q}} \sum_{t=1}^{n_{q}} \nu(\tilde{x}y)^{q}_{st}.$$  

As in (2.21) we get a global section of the bundle system $(\tilde{E}, \tilde{V})$,

$$\partial_{I, H} = \left\{ (\partial_{I, H}^{(0)}_{I_{1}\times I_{2}})_{I_{1}\times I_{2}} = \partial_{I, H}^{(0)} + (\nu(\tilde{x}y)^{q})_{I_{1}\times I_{2}} \mid I_{1} \times I_{2} \in \mathcal{N}^{0}(\tilde{x}, \tilde{y}) \right\}.$$  

Then it easily follows from (2.29) and the choice of this section that

$$\mathcal{M}_{I_{1}}^{\nu(\tilde{x}y)} \oplus \mathcal{M}_{I_{2}}^{\nu(\tilde{z}q)} \subset \mathcal{M}_{I_{1}\times I_{2}}^{\nu(\tilde{x}y)} \subset \left( (\partial_{I, H}^{(0)})_{I_{1}\times I_{2}} \right)^{-1}(0).$$
for $I_1 \times I_2 \in \mathcal{N}(\bar{x}, \bar{z}_q) \times \mathcal{N}(\bar{z}_q, \bar{y})$, and that each section \((\bar{g}^\nu_{J,H})_{I_1 \times I_2})\) is transversal to the zero section at all points of 
\[
\tilde{\mathcal{M}}_{I_1}^{\nu(\bar{x}, \bar{z}_q)} \# \tilde{\mathcal{M}}_{I_2}^{\nu(\bar{z}_q, \bar{y})} \subset \tilde{W}_{I_1}^{\nu(\bar{x}, \bar{z}_q)} \# \tilde{W}_{I_2}^{\nu(\bar{z}_q, \bar{y})}.
\]

The last set consists of all points \((u_s \# v_t)_{(s,t) \in I_1 \times I_2}\) in 
\[
\prod_{(s,t) \in I_1 \times I_2} \tilde{W}(f^{(1)}_{s,t}) \# \tilde{W}(f^{(2)}_{s,t})
\]
such that \(\pi_{s,t}^q (u_s \# v_t) = \pi_{s',t'}^q (u_{s'} \# v_{t'})\) for any \((s, t)\) and \((s', t')\) in \(I_1 \times I_2\).

By the stability of a surjective map under small perturbation we can show that there exist \(\Gamma_{I_1 \times I_2}^{(q)}\)-invariant open neighborhoods of \(\tilde{\mathcal{M}}_{I_1}^{\nu(\bar{x}, \bar{z}_q)} \# \tilde{\mathcal{M}}_{I_2}^{\nu(\bar{z}_q, \bar{y})}\) in \(\tilde{V}_{I_1 \times I_2}\),

\[
(2.32)
\]
\[
\tilde{V}_{I_1 \times I_2}^0 \subset \subset \tilde{V}_{I_1 \times I_2}^1
\]
such that the section \((\bar{g}^\nu_{J,H})_{I_1 \times I_2})\) is transversal to the zero section at points of

\[
(2.33)
\]
\[
\tilde{V}_{I_1 \times I_2}^1 \cap \tilde{\mathcal{M}}_{I_1 \times I_2}^{\nu(\bar{x}, \bar{z}_q)} \# \tilde{\mathcal{M}}_{I_2}^{\nu(\bar{z}_q, \bar{y})}.
\]

So the space in (2.33) is a cornered and stratified Banach variety that has the dimension given by the Index Theorem on all of its strata. Moreover, the collection

\[
\left\{ \tilde{\mathcal{M}}_{I_1 \times I_2}^{\nu(\bar{x}, \bar{z}_q)} \mid I_1 \times I_2 \in \mathcal{N}(\bar{x}, \bar{z}_q) \times \mathcal{N}(\bar{z}_q, \bar{y}) \right\}
\]
also satisfy the compatibility as in Theorem 2.6. In particular the open neighborhoods in (2.32) can be chosen to guarantee that the collection

\[
\left\{ \tilde{V}_{I_1 \times I_2}^1 \cap \tilde{\mathcal{M}}_{I_1 \times I_2}^{\nu(\bar{x}, \bar{z}_q)} \mid I_1 \times I_2 \in \mathcal{N}(\bar{x}, \bar{z}_q) \times \mathcal{N}(\bar{z}_q, \bar{y}) \right\}
\]
satisfies the compatibility. Note that the projection image of \(\tilde{V}_{I_1 \times I_2}^1 \cap \tilde{\mathcal{M}}_{I_1 \times I_2}^{\nu(\bar{x}, \bar{z}_q)}\) in \(B(\bar{x}, \bar{y})\) is not compact in general (unlike in Theorem 2.6). We denote by

\[
\left( \tilde{V}_{I_1 \times I_2}^1 \cap \tilde{\mathcal{M}}_{I_1 \times I_2}^{\nu(\bar{x}, \bar{z}_q)} \right)^c,
\]

the union of the top and 1-codimensional strata of \(\tilde{V}_{I_1 \times I_2}^1 \cap \tilde{\mathcal{M}}_{I_1 \times I_2}^{\nu(\bar{x}, \bar{z}_q)}\), then it is a smooth manifold with boundary of dimension \(\mu(\bar{x}) - \mu(\bar{y}) - 1\) that is contained in the smooth locus of \(\tilde{V}_{I_1 \times I_2}^1 \cap \tilde{\mathcal{M}}_{I_1 \times I_2}^{\nu(\bar{x}, \bar{z}_q)}\) and that has the boundary

\[
\partial \left( \tilde{V}_{I_1 \times I_2}^1 \cap \tilde{\mathcal{M}}_{I_1 \times I_2}^{\nu(\bar{x}, \bar{z}_q)} \right)^c = \tilde{\mathcal{M}}_{I_1 \times I_2}^{\nu(\bar{x}, \bar{z}_q), \# \tilde{\mathcal{M}}_{I_2}^{\nu(\bar{z}_q, \bar{y})}, \partial}
\]
if \(I_1 \times I_2 \in \mathcal{N}(\bar{x}, \bar{z}_q) \times \mathcal{N}(\bar{z}_q, \bar{y})\). In the following we shall extend

\[
\left\{ \tilde{V}_{I_1 \times I_2}^1 \cap \tilde{\mathcal{M}}_{I_1 \times I_2}^{\nu(\bar{x}, \bar{z}_q)} \mid I_1 \times I_2 \in \mathcal{N}(\bar{x}, \bar{y}) \right\}
\]
into a virtual moduli cycle for $\overline{\mathcal{M}}(\bar{x}, \bar{y}; J, H)$ under the condition that
\[
\left\{ \overline{V}_{I_1 \times I_2}^0 \cap \overline{\mathcal{M}}^{\nu(\bar{x} \bar{y})_0} | I_1 \times I_2 \in \mathcal{N}(\bar{x}, \bar{y})^0 \right\}
\]
is not changed. Since both $\overline{\mathcal{M}}(\bar{x}, \bar{y}; J, H)$ and
\[
\bigcup_{q=1}^r \overline{\mathcal{M}}(\bar{x}, z_q; J, H) \# \overline{\mathcal{M}}(z_q, \bar{y}; J, H)
\]
are compact we can take points $(\tilde{\gamma}, \ldots, \tilde{\gamma})$ satisfying the conditions for the construction of the virtual moduli cycle above. Let us choose $\Gamma_{h_j}$-invariant smooth cut-off functions $\gamma(h_j)$ on $\overline{W}(h_j)$ such that
\[
\pi_j(\text{supp}(\gamma(h_j))) \cap V_{I_1 \times I_2}^0 = \emptyset \ \forall 1 \leq j \leq n, I_1 \times I_2 \in \mathcal{N}(\bar{x}, \bar{y})^0
\]
where $V_{I_1 \times I_2}^0$ is the projection image of $\overline{V}_{I_1 \times I_2}^0$ in $\mathcal{B}(\bar{x}, \bar{y})$.

Let $(\overline{E}(\bar{x}\bar{y}), \overline{V}(\bar{x}\bar{y}))$ and $\mathcal{N}(\bar{x}, \bar{y})$ be the corresponding bundle system and index set to this covering. Assume that
\[
\nu(\bar{x} \bar{y})_st^q = \left\{ \left( \nu(\bar{x} \bar{y})_st^q \right)_I | I \in \mathcal{N}(\bar{x}, \bar{y}) \right\}
\]
is a global section of this bundle system obtained by the cut-off function $\gamma(f_{st}^q)$ and $\nu(\bar{x} \bar{y})_st^q$ as in Step 2. Of course, for each $\nu_j \in R_\epsilon(h_j)$ we still denote by $\overline{\nu}_j = \{(\overline{\nu}_j)_I | I \in \mathcal{N}(\bar{x}, \bar{y})\}$ the global section of $(\overline{E}(\bar{x}\bar{y}), \overline{V}(\bar{x}\bar{y}))$ obtained from $\gamma(h_j)$ and $\nu_j$ as below (2.18).

For $\delta > 0$ we assume that $Z_{\delta}^\ast(\{h_j\})$ is a $\delta$-neighborhood of zero of $\oplus_{j=1}^n R_\epsilon(h_j)$. As before we have a bundle system $(\overline{E}(\bar{x}\bar{y}) \times Z_{\delta}^\ast(\{h_j\}), \overline{V}(\bar{x}\bar{y}) \times Z_{\delta}^\ast(\{h_j\}))$ and a well-defined global section of it
\[
\overline{\partial}_{J, H} + \hat{\epsilon} : (u_I, \nu) \mapsto \overline{\partial}_{J, H} u_I + \sum_{j=1}^{n} (\overline{\nu}_j)_I(u_I) + \sum_{q=1}^{r} \sum_{s=1}^{m_q} \sum_{t=1}^{n_q} \left( \nu(\bar{x} \bar{y})_st \right)_I(u_I)
\]
for $(u_I, \nu) = (u_I, (\nu_1, \ldots, \nu_n)) \in \overline{V}(\bar{x}\bar{y}) \times Z_{\delta}^\ast(\{h_j\})$. Notice that our choices above imply this section to be transversal to the zero section for $\delta > 0$ small enough. It follows that for a generic choice of $\nu \in Z_{\delta}^{\ast}(\{h_j\})$ the section of $(\overline{E}(\bar{x}\bar{y}), \overline{V}(\bar{x}\bar{y}))$ given by
\[
(2.34) \quad u_I \mapsto \overline{\partial}_{J, H} u_I + \sum_{j=1}^{n} (\overline{\nu}_j)_I(u_I) + \sum_{q=1}^{r} \sum_{s=1}^{m_q} \sum_{t=1}^{n_q} \left( \nu(\bar{x} \bar{y})_st \right)_I(u_I)
\]
for $u_I \in \overline{V}(\bar{x}\bar{y})_I$, is transversal to the zero section.
Setting $\nu(\vec{x}\vec{y}) := (\oplus_{q=1}^{r} \oplus_{s=1}^{m_q} \oplus_{t=1}^{n_q} \nu(\vec{x}\vec{y})^q_{st}) \oplus (\oplus_{j=1}^{n_j} \nu_j)$ then it belongs to $(\oplus_{q=1}^{r} \oplus_{s=1}^{m_q} \oplus_{t=1}^{n_q} R_\mu(f_s^{(1 q)}) \oplus R_\mu(f_t^{(2 q)})) \oplus (\oplus_{j=1}^{n_j} R_\mu(h_j))$, and the section
\[
\widehat{\mathcal{J}}_{\nu(\vec{x}\vec{y})} = \left\{ \left( \widehat{\mathcal{J}}_{\nu(\vec{x}\vec{y})}^I \right)_I \middle| I \in \mathcal{N}(\vec{x}, \vec{y}) \right\}
\]
of $(\vec{E}(\vec{x}\vec{y}), \vec{V}(\vec{x}\vec{y}))$ is transversal to the zero section. Here $\left( \widehat{\mathcal{J}}_{\nu(\vec{x}\vec{y})}^I \right)_I$ is given by (2.34). Denote by $C\left( \mathcal{M}^{\nu(\vec{x}\vec{y})}(\vec{x}, \vec{y}) \right)$ the virtual moduli cycle constructed from this section. It is easy to see that its boundary is given by
\[
\sum_{q=1}^{r} \sum_{I_1 \in \mathcal{N}(\vec{x}, \vec{z}_q)} \sum_{I_2 \in \mathcal{N}(\vec{z}_q, \vec{y})} \frac{1}{|\Gamma_{I_1}^{(1 q)}| \cdot |\Gamma_{I_2}^{(2 q)}|} \cdot \mathcal{M}^{\nu(\vec{z}_q),\nu(\vec{z}_q),\nu(\vec{z}_q)}_{I_1} \times \mathcal{M}^{\nu(\vec{z}_q),\nu(\vec{z}_q),\nu(\vec{z}_q)}_{I_2},
\]
which can be identified with
\[
\sum_{q=1}^{r} \sum_{I_1 \in \mathcal{N}(\vec{x}, \vec{z}_q)} \sum_{I_2 \in \mathcal{N}(\vec{z}_q, \vec{y})} \frac{1}{|\Gamma_{I_1}^{(1 q)}| \cdot |\Gamma_{I_2}^{(2 q)}|} \cdot \mathcal{M}^{\nu(\vec{z}_q),\nu(\vec{z}_q),\nu(\vec{z}_q)}_{I_1} \times \mathcal{M}^{\nu(\vec{z}_q),\nu(\vec{z}_q),\nu(\vec{z}_q)}_{I_2}.
\]
But the latter is just $\sum_{q=1}^{r} C\left( \mathcal{M}^{\nu(\vec{z}_q)}(\vec{x}, \vec{z}_q) \right) \times C\left( \mathcal{M}^{\nu(\vec{z}_q)}(\vec{z}_q, \vec{y}) \right)$. (2.12) is proved.

Now we first construct all relative virtual moduli cycles $C\left( \mathcal{M}^{\nu(\vec{x}\vec{y})}(\vec{x}, \vec{y}) \right)$ for all $\vec{x}, \vec{y} \in \vec{P}(H)$ with $\mu(\vec{x}) - \mu(\vec{y}) = 1$, such that
\[
(2.35) \quad \sharp(\mathcal{M}^{\nu}(\vec{x}, \vec{y})) = \sharp(\mathcal{M}^{\nu}(\vec{x}, -A), \vec{y}, -A))
\]
for any $A \in \Gamma$. Then we follow the above methods to construct all relative virtual moduli cycles $C\left( \mathcal{M}^{\nu(\vec{z}_q)}(\vec{x}, \vec{z}_q) \right)$ for all $\vec{x}, \vec{y} \in \vec{P}(H)$ with $\mu(\vec{x}) - \mu(\vec{y}) = 2$. The family of the relative virtual moduli cycles are compatible and thus satisfy the requirements in the definition of $\partial^F$.

Remark 2.7. As in [LiuT2] and [LiuT3] we need the virtual moduli cycles of dimension more than one in this paper, and can use the notion of local components in [LiuT3] to construct a desingularization of the bundle system $(\vec{E}^I, \vec{W}^I)$, a new bundle system $(\vec{E}^I, \vec{W}^I) = \left\{ (\vec{E}^I_I, \vec{W}^I_I) \middle| I \in \mathcal{N}(\vec{x}, \vec{y}) \right\}$ such that each $\vec{E}^I_I$ (resp. $\vec{W}^I_I$) is a stratified Banach manifold (resp. bundle). Then replacing $(\vec{E}^I, \vec{W}^I)$ everywhere by $(\vec{E}^I_I, \vec{W}^I_I)$ in the previous construction of the virtual moduli cycles one can get a virtual moduli cycle
\[
\sum_{I \in \mathcal{N}(\vec{x}, \vec{y})} \frac{1}{|\Gamma_I|} \left\{ \pi_I : \mathcal{M}^\nu_I \rightarrow W \right\}, \text{ still denoted by } \mathcal{M}^{\nu}(\vec{x}, \vec{y}),
\]
such that each $\mathcal{M}^\nu_I$ is a cornered smooth manifold. We can use the same method to make suitable modifications for the above arguments and in the
case \( \mu(\bar{x}) - \mu(\bar{y}) = 2 \) obtain the conclusion corresponding with (2.12), i.e.,

\[
\partial \mathcal{M}^{\nu(\bar{z}\bar{y})}(\bar{x}, \bar{y}) = \sum_{\mu(\bar{z}) = \mu(\bar{x}) - 1} \mathcal{M}^{\nu(\bar{z}\bar{x})}(\bar{x}, \bar{z}) \times \mathcal{M}^{\nu(\bar{z}\bar{y})}(\bar{z}, \bar{y}).
\]

Hence we can replace \( \sharp C \left( \mathcal{M}^{\nu(\bar{z}\bar{y})}(\bar{x}, \bar{y}) \right) \) by \( \sharp \mathcal{M}^{\nu(\bar{z}\bar{y})}(\bar{x}, \bar{y}) \) in (1.2).

Now we begin to construct all virtual moduli cycles \( \mathcal{M}^{\nu}(\bar{x}; H, J) \) in (2.9) and (2.10) which are compatible with the relative virtual cycles used in (1.2). We only consider \( \mathcal{M}^{\nu}_+(\bar{x}; H, J) \). Denote by \( T\mathcal{M}^{\nu}_+(\bar{x}; H, J) \) and \( B\mathcal{M}^{\nu}_+(\bar{x}; H, J) \) its top strata and 1-codimensional strata. If \( \dim \mathcal{M}^{\nu}_+(\bar{x}; H, J) = \dim M - \mu(\bar{x}) > 0 \), then any element \( f_+ \) of \( B\mathcal{M}^{\nu}_+(\bar{x}; H, J) \) must have a form \( f_+ = (h_+, g) \). Here \( g \in \mathcal{M}^{\nu_2}(\bar{y}; \bar{x}; H, J) \) and \( \mu(\bar{y}) - \mu(\bar{x}) = 1 \), and \( h_+ \in T\mathcal{M}^{\nu_1}_+(\bar{y}; H, J) \). So we can construct each \( \mathcal{M}^{\nu}_+(\bar{x}; H, J) \) inductively with respect to \( \dim M - \mu(\bar{x}) \). In fact, if these have been constructed for \( \dim M - \mu(\bar{x}) = 0 \), then for each \( \bar{x} \in \bar{\mathcal{P}}(H) \) with \( \dim M - \mu(\bar{x}) = 1 \), we can use them and \( \mathcal{M}^{\nu}(\bar{y}, \bar{x}; H, J) \) used in (1.2) to construct

\[
B\mathcal{M}^{\nu}_+(\bar{x}; H, J) := \bigcup_{\mu(\bar{y}) = \mu(\bar{x}) - 1} T\mathcal{M}^{\nu}_+(\bar{y}; H, J) \times \mathcal{M}^{\nu}(\bar{y}, \bar{x}; H, J).
\]

Next as done in the Proof of (2.12) we can extend \( B\mathcal{M}^{\nu}_+(\bar{x}; H, J) \) into a derived family \( \mathcal{M}^{\nu}_+(\bar{x}; H, J) = \{ \mathcal{M}^{\nu}_+(\bar{x}; H, J) \mid I \in \mathcal{N}^+(\bar{x}) \} \) for a virtual moduli cycle \( \mathcal{M}^{\nu}_+(\bar{x}; H, J) \). Let \( \{ \mathcal{M}^{\nu}_+(\bar{x}; H, J) \mid \bar{x} \in \bar{\mathcal{P}}(H) \} \) be all virtual moduli cycles constructed by induction. Then it holds that

\[
B\mathcal{M}^{\nu}_+(\bar{x}; H, J) = \bigcup_{\mu(\bar{y}) = \mu(\bar{x}) - 1} T\mathcal{M}^{\nu}_+(\bar{y}; H, J) \times \mathcal{M}^{\nu}(\bar{y}, \bar{x}; H, J).
\]

Consider the natural orientations on them again we can write

\[
2.36 \quad B\mathcal{M}^{\nu}_+(\bar{x}; H, J) = \sum_{\mu(\bar{y}) = \mu(\bar{x}) + 1} \sharp \left( C(\mathcal{M}^{\nu}(\bar{y}, \bar{x})) \right) \cdot T\mathcal{M}^{\nu}_+(\bar{y}; H, J).
\]

Here the rational numbers \( \sharp \left( C(\mathcal{M}^{\nu}(\bar{y}, \bar{x})) \right) \) is as in (1.2), and the identity in (2.36) is understood as follows: Since the sums at the right side of (2.36) are finite we can take \( L > 0 \) to be the smallest common multiple of denominators of all rational numbers \( \sharp \left( C(\mathcal{M}^{\nu}(\bar{y}, \bar{x})) \right) \). Then (2.36) is equivalent to

\[
L \cdot B\mathcal{M}^{\nu}_+(\bar{x}; H, J) = \sum_{\mu(\bar{y}) = \mu(\bar{x}) + 1} \left( L \cdot \sharp \left( C(\mathcal{M}^{\nu}(\bar{y}, \bar{x})) \right) \right) \cdot T\mathcal{M}^{\nu}_+(\bar{y}; H, J).
\]

For this identity the left side is understood as the disjoint union of \( L \) copies of \( B\mathcal{M}^{\nu}_+(\bar{x}; H, J) \), and the right side is also the disjoint union that contains, for each \( \bar{y} = [y, u] \) with \( \mu(\bar{y}) = \mu(\bar{x}) + 1 \), \( L \cdot \sharp \left( C(\mathcal{M}^{\nu}(\bar{y}, \bar{x})) \right) \) copies of \( T\mathcal{M}^{\nu}_+(\bar{y}; H, J) \) as \( \sharp \left( C(\mathcal{M}^{\nu}(\bar{y}, \bar{x})) \right) > 0 \), and \( (-L) \cdot \sharp \left( C(\mathcal{M}^{\nu}(\bar{y}, \bar{x})) \right) \) copies of \( B\mathcal{M}^{\nu}_+(\bar{y}; H, J) \) as \( \sharp \left( C(\mathcal{M}^{\nu}(\bar{y}, \bar{x})) \right) = 0 \).
copies of \( T\mathcal{M}_+^+(\overline{y}; H, J)^* \) as \( \sharp(C(\mathcal{M}^+ (\overline{y}, \overline{x})) < 0 \), where \( T\mathcal{M}_+^+(\overline{y}; H, J)^* \) is \( T\mathcal{M}_+^+(\overline{y}; H, J) \) with the orientation reversed. In other words, it is understood as the topological sum. With the same methods we can construct the compatible \( \mathcal{M}^- (\overline{x}; H, J) \) and obtain

\[
B\mathcal{M}^+_-(\overline{x}; H, J) = \sum_{\mu(\overline{x})=\mu(\overline{z})-1} \sharp(C(\mathcal{M}^+ (\overline{x}, \overline{z})) \cdot T\mathcal{M}^- (\overline{z}; H, J). \tag{2.37}
\]

### 3. The intersections of virtual moduli cycles with stable and unstable manifolds.

For the materials on Morse homology the readers may refer to [AuB, Sch1] and [Sch4]. Given a Morse-Smale pair \((h, g)\) on \( M \) with \( h \in O(h_0) \) the stable and unstable manifolds of a critical point \( a \in \text{Crit}(h) \) are given by

\[
W^s(a, h, g) := \{ \gamma : [0, \infty) \rightarrow M | \dot{\gamma}(s) + \nabla g h(\gamma(s)) = 0, \gamma(\infty) = a \},
\]

\[
W^u(a, h, g) := \{ \gamma : (-\infty, 0] \rightarrow M | \dot{\gamma}(s) + \nabla g h(\gamma(s)) = 0, \gamma(-\infty) = a \}.
\]

There are two obvious evaluations

\[
E^s_a : W^s(a, h, g) \rightarrow M, \gamma \mapsto \gamma(0) \quad \text{and} \quad E^u_a : W^u(a, h, g) \rightarrow M, \gamma \mapsto \gamma(0).
\]

Both are also smooth embeddings into \( M \). Throughout this paper we fix orientations for all unstable manifolds \( W^u(a, h, g) \), then the orientation of \( M \) induces orientations for \( W^s(a, h, g) \) and \( W^u(a, h, g) \cap W^s(b, h, g) \). We wish to study the intersection numbers of the maps in (2.11) and (3.1) under some conditions. As usual we need their compactifications of the following versions. Consider the disjoint union \( \overline{W}^u(a, h, g) := W^u(a, h, g) \cup S\overline{W}^u(a, h, g) \), where \( S\overline{W}^u(a, h, g) \) is the disjoint unions

\[
\bigcup_{\gamma(0)} \mathcal{M}_{a_i - a_i} (h, g) \times \cdots \times \mathcal{M}_{a_1} (h, g) \times W^u(a_i, h, g)
\]

for all critical points \( a_0 = a, \ldots, a_i \) with the Morse indexes \( \mu(a_0) > \cdots > \mu(a_i) \). Similarly, in the disjoint union \( \overline{W}^s(a, h, g) := W^s(a, h, g) \cup S\overline{W}^s(a, h, g) \), the second set \( S\overline{W}^s(a, h, g) \) is the disjoint unions

\[
\bigcup_{\gamma(0)} W^s(a_i, h, g) \times \mathcal{M}_{a_i - a_i} (h, g) \times \cdots \times \mathcal{M}_{a_1} (h, g)
\]

for critical points \( a_0 = a, \ldots, a_i \) with the Morse indexes \( \mu(a_0) < \cdots < \mu(a_i) \).

As usual the compactness and gluing arguments in Morse homology (see [AuB] and [Sch1]) give:

**Lemma 3.1.** The sets \( \overline{W}^u(a, h, g) \) and \( \overline{W}^s(a, h, g) \) may be topologized with a natural way so that they are the compactifications of \( W^u(a, h, g) \) and \( W^s(a, h, g) \) respectively, and that \( \partial \overline{W}^u(a, h, g) = S\overline{W}^u(a, h, g) \) and \( \partial \overline{W}^s(a, h, g) = S\overline{W}^s(a, h, g) \). Moreover these compactified spaces both have the
structure of a manifold with corners, and maps $E^u_a$ and $E^s_a$ may smoothly extend to them, denoted by $E^u_a$ and $E^s_a$, which also give natural injective immersions from these two compactified spaces into $M$.

The following lemma, slightly different from Theorem 4.9 in [Sch4], may be easily proved (cf. [Lu2]).

**Lemma 3.2.** Let $R$ be the set of all Riemannian metrics on $M$. Then for any smooth map $\chi$ from a smooth manifold $V$ to $M$ there exists a Baire subset $(O \times R)_{reg} \subset O(h_0) \times R$ such that for each pair $(h, g) \in (O(h_0) \times R)_{reg}$ and $a \in \text{Crit}(h)$ the maps $E^u_a$ and $E^s_a$ are transverse to $\chi$. Consequently, the spaces

$$
M^s_{\chi,a}(h,g) := \{(p, \gamma) \in V \times W^s(a,h,g) \mid \chi(p) = E^s_a(\gamma) = \gamma(0)\},
$$

$$
M^u_{\chi,a}(h,g) := \{(p, \gamma) \in V \times W^u(a,h,g) \mid \chi(p) = E^u_a(\gamma) = \gamma(0)\}
$$

are respectively smooth manifolds of $\dim V - \mu(a)$ and $\dim V + \mu(a) - 2n$.

As usual, if $V, M, W^s(a,h,g)$ and $W^u(a,h,g)$ are oriented then $M^s_{\chi,a}(h,g)$ and $M^u_{\chi,a}(h,g)$ have the natural induced orientations. Specially, if $M^s_{\chi,a}(h,g)$ and $M^u_{\chi,a}(h,g)$ are of dimension zero we have the oriented intersection numbers $\chi \cdot E^s_a$ and $\chi \cdot E^u_a$ which counts the algebraic sum of the oriented points in $M^s_{\chi,a}(h,g)$ and $M^u_{\chi,a}(h,g)$ respectively. Note that the 1-codimensional stratum of $W^u(a,h,g)$ (resp. $W^s(a,h,g)$) is given by

$$
\sum_{\mu(b) = \mu(a) - 1} n(a,b) \cdot W^u(b,h,g) \quad \text{(resp.} \quad \sum_{\mu(c) = \mu(a) + 1} n(c,a) \cdot W^s(c,h,g)\text{)},
$$

where $n(a,b)$ and $n(c,a)$ are as in (1.6).

Since there are only countable manifolds involved in our arguments using Claim A.1.11 in [LO] we can always fix an $h \in O(h_0)$ such that all transversal arguments hold for a generic Riemannian metric $g$ on $M$. By Lemmas 3.1 and 3.2 and these remarks we immediately obtain:

**Proposition 3.3.** If $\mu(a) = \mu(\bar{x})$ then for a generic Riemannian metric $g$ on $M$ the maps $E^u_a$ and $E^s_a \cdot EV^u_+(\bar{x})$, and $E^s_a$ and $E^s_a \cdot EV^-_-(\bar{x})$ are transversal, and their intersection numbers are the well-defined rational numbers $E^s_a \cdot EV^s_+(\bar{x})$ and $E^s_a \cdot EV^-_-(\bar{x})$.

From now on, for the sake of clearness we denote by

$$
n^+_{\nu^+}(a, \bar{x}) \equiv n^+_{\nu^+}(a, \bar{x}, H, J; h, g) := E^u_a \cdot EV^u_+(\bar{x}),
$$

$$
n^-_{\nu^-}(a, \bar{x}) \equiv n^-_{\nu^-}(a, \bar{x}, H, J; h, g) := E^s_a \cdot EV^-_-(\bar{x}).
$$

The standard arguments show that these numbers are independent of small regular $\nu^\pm$. Later, we no longer state this and often omit $\nu^\pm$ without occurrence of confusion. It easily follows from (2.7) that:
Proposition 3.4. If $n_\pm(a, [x,v]) \neq 0$ then $\pm \int_{D^2} v^* \omega \geq -2 \max |H|.$

Using (2.36), (2.37) and Lemmas 3.1 and 3.2 we can obtain the following two results:

Proposition 3.5. If $\mu(a) - \mu(\tilde{x}) = 1$ then for a generic Riemannian metric $g$ on $M$ the fibre product

$$\overline{W}^\mu(a, h, g) \times_{E_g = \text{EV}^\mu_+ (\tilde{x})} \overline{M}^\mu_+(\tilde{x}; H, J)$$

is still a collection of compatible local cornered smooth manifolds of dimension 1 and with the natural orientations. Its boundary is given by

$$\left( \bigcup_{\mu(b) = \mu(a) - 1} n(a, b) \cdot \left( W^\mu(a, h, g) \times_{E_g = \text{EV}^\mu_+ (\tilde{x})} \overline{M}^\mu_+(\tilde{x}; H, J) \right) \right) \cup$$

$$\left( - \bigcup_{\mu(\tilde{v}) = \mu(\tilde{x}) + 1} \sharp (C(\overline{M}^\nu(\tilde{y}, \tilde{x}))) \cdot \left( W^\mu(a, h, g) \times_{E_g = \text{EV}^\nu_+ (\tilde{y})} \overline{M}^\nu_+(\tilde{y}; H, J) \right) \right).$$

Notice that the projection onto $M \times W_+ (\tilde{x}, H, J)$ of this set is a finite set. Let us denote by $\sharp (\partial ( W^\mu(a, h, g) \times_{E_g = \text{EV}^\nu_+ (\tilde{x})} \overline{M}^\nu_+(\tilde{x}; H, J)))$ the number of elements of this finite set counted with appropriate signs and rational weights. Then this number must be zero, and it follows that

$$\sum_{\mu(b) = \mu(a) - 1} n(a, b) \cdot n_\nu^+(b, \tilde{x}) = \sum_{\mu(\tilde{v}) = \mu(\tilde{x}) + 1} n_\nu^+(a, \tilde{v}) \cdot \sharp (C(\overline{M}^\nu(\tilde{y}, \tilde{x})))$$

Remark that the fibre product in Proposition 3.5 has boundary. But we assume its dimension to be 1. Hence its boundary agrees with the 1-codimensional stratum of the fibre product. This remark is still valid for the following proposition:

Proposition 3.6. If $\mu(\tilde{x}) - \mu(a) = 1$ then for a generic Riemannian metric $g$ on $M$ the fibre product

$$\overline{W}^\nu(a, h, g) \times_{E_g = \text{EV}^\nu_-(\tilde{x})} \overline{M}^\nu_-(\tilde{x}; H, J)$$

is a collection of compatible local cornered smooth manifolds of dimension 1 and with the natural orientations. Its boundary is given by

$$\left( \bigcup_{\mu(a) = \mu(b) - 1} n(b, a) \cdot \left( W^\nu(b, h, g) \times_{E_g = \text{EV}^\nu_- (\tilde{x})} \overline{M}^\nu_- (\tilde{x}; H, J) \right) \right) \cup$$

$$\left( - \bigcup_{\mu(\tilde{z}) = \mu(\tilde{x}) - 1} \sharp (C(\overline{M}^\nu(\tilde{x}, \tilde{z}))) \cdot \left( W^\nu(a, h, g) \times_{E_g = \text{EV}^\nu_- (\tilde{z})} \overline{M}^\nu_- (\tilde{z}; H, J) \right) \right).$$

Consequently, $\sharp (\partial ( W^\nu(a, h, g) \times_{E_g = \text{EV}^\nu_- (\tilde{x})} \overline{M}^\nu_- (\tilde{x}; H, J))) = 0$ implies that

$$\sum_{\mu(\tilde{z}) = \mu(\tilde{x}) - 1} \sharp (C(\overline{M}^\nu(\tilde{x}, \tilde{z}))) \cdot n^-_\nu(a, \tilde{z}) = \sum_{\mu(b) = \mu(a) + 1} n^-_\nu(b, \tilde{x}) \cdot n(b, a).$$
Remark 3.7. Using Proposition 3.4 it is easily proved that $\Phi$ and $\Psi$ are $\Lambda_\omega$-module homomorphisms (see [Lu2]). Moreover, using Propositions 3.4, 3.5, 3.6 and (2.35) we can prove that $\Phi \circ \delta_k^Q = \delta_k^F \circ \Phi$ and $\Psi \circ \delta_k^F = \delta_k^Q \circ \Psi$ for every integer $k$. That is, $\Phi$ and $\Psi$ also induce the homomorphisms between two corresponding homology groups (see [Lu2]).

4. Proof of Theorem 1.1.

Theorem 1.1 may follow from the following Theorems 4.1 and 4.9 immediately.

Theorem 4.1. $\Psi \circ \Phi$ is chain homotopy equivalent to the identity. Consequently, $\Phi$ induces an injective $\Lambda_\omega$-module homomorphism from $QH_*(h,g;\mathbb{Q})$ to $HF_*(M,\omega;H,J,\nu;\mathbb{Q})$.

Proof. We shall prove it in four steps.

Step 1. For every $\langle a,A \rangle \in \text{Crit}(h) \times \Gamma$ with $\mu(\langle a,A \rangle) = k$ we have

$$\Psi \circ \Phi(\langle a,A \rangle) = \sum_{\mu(b,B) = k} m^\nu_{+,-}(\langle a,A \rangle;\langle b,B \rangle)\langle b,B \rangle,$$

where $m^\nu_{+,-}(\langle a,A \rangle;\langle b,B \rangle)$ is given by

$$\sum_{\mu(\tilde{x}) = k} n^\nu_{+}(a,\tilde{x}_a(-A)) \cdot n^\nu_{-}(b,\tilde{x}_b(-B)). \quad (4.1)$$

Notice here that $\langle a,A \rangle$ and $\langle b,B \rangle$ satisfy

$$\mu(a) - \mu(b) + 2c_1(B - A) = 0. \quad (4.2)$$

Firstly, we claim that the sum in (4.1) is finite. In fact, if $\tilde{x} \in \tilde{P}_k(H)$ is such that

$$n^\nu_{+}(a,\tilde{x}_a(-A)) \cdot n^\nu_{-}(b,\tilde{x}_b(-B)) \neq 0, \quad (4.3)$$

by Proposition 3.4 we have

$$\omega(A) - 2\max|H| \leq \int_{D^2} v^*\omega \leq \omega(B) + 2\max|H|. \quad (4.4)$$

It easily follows that the number of such $\tilde{x} \in \tilde{P}(H)$ is finite. Let them be $\tilde{x}_1,\ldots,\tilde{x}_s$. Then for a generic $(h,g) \in O(h_0) \times \mathcal{R}$, it holds

$$m^\nu_{+,-}(\langle a,A \rangle;\langle b,B \rangle) = \sum_{i=1}^{s} n^\nu_{+}(a,\tilde{x}_i^a(-A)) \cdot n^\nu_{-}(b,\tilde{x}_i^b(-B)). \quad (4.5)$$

Note that for each $\tilde{x} \in \tilde{P}_k(H)$ the product $n^\nu_{+}(a,\tilde{x}_a^a(-A)) \cdot n^\nu_{-}(b,\tilde{x}_b^a(-B))$ can be explained as the rational intersection number

$$(E^u_a \times E^b_b) \cdot (E^u_{+}(\tilde{x}_a(-A)) \times E^v_{-}(\tilde{x}_b(-B))). \quad (4.6)$$
of the product map
\[ \text{EV}^\nu_+ (\tilde{x}(A)) \times \text{EV}^\nu_- (\tilde{x}(B)) \]
from \( \mathcal{M}^\nu_+ (\tilde{x}(A); H, J) \times \mathcal{M}^\nu_- (\tilde{x}(B); H, J) \) to \( M \times M \) and the evaluation
\[ (4.7) \quad E^u_a \times E^s_b : W^u(a, h, g) \times W^s(b, h, g) \to M \times M \]
for a generic \((h, g) \in \mathcal{O}(h_0) \times \mathcal{R}\). But the intersection number in (4.6) is the number of the oriented points with rational weights in the fibre product
\[ (4.8) \quad (W^u(a, h, g) \times W^s(b, h, g)) \times_R (\mathcal{M}^\nu_+(\tilde{x}(A)) \times \mathcal{M}^\nu_- (\tilde{x}(B))). \]
Here \( R \) is representing \( \text{EV}^\nu_+ \times \text{EV}^\nu_- = E^u_a \times E^s_b \), and we have omitted \( H, J \)
in \( \mathcal{M}^\nu_\pm (\cdot; H, J) \). However, because of dimension relations the fibre product in (4.8) is an empty set for a generic \((h, g) \in \mathcal{O}(h_0) \times \mathcal{R}\) if \( \mu([x, v]) \neq k \). As usual we still understand the intersection number being zero in this case. Hence (4.1) and (4.5) become
\[ (4.9) \quad m_{\nu, -}(\langle a, A \rangle; \langle b, B \rangle) = \sum_{\tilde{z} \in \tilde{P}(H)} (E^u_a \times E^s_b) \cdot (\text{EV}^\nu_+ (\tilde{x}(A)) \times \text{EV}^\nu_- (\tilde{x}(B))). \]

**Step 2.** We need to give another explanation of the intersection number in (4.6). To this goal we note that for a given \( D \in \Gamma \) a positive disk \( u_+ \in \mathcal{M}_+ (\tilde{y}; H, J) \) and a negative disk \( u_- \in \mathcal{M}_-(\tilde{y}_\nu(-D); H, J) \) may be glued into a sphere in \( M \) along \( y \), denoted by \( u_+#u_- \). Using (2.2) and (2.4) one easily checks that it is a representative of \( D \). Let us denote by \( \mathcal{M}_D (\tilde{y}; H, J) \) by the space of all such \( u_+#u_- \). Clearly, it can be identified with the product space \( \mathcal{M}_+ (\tilde{y}; H, J) \times \mathcal{M}_-(\tilde{y}_\nu(-D); H, J) \). Therefore, its virtual dimension is equal to \( \dim M + 2c_1(D) \). We denote by
\[ \mathcal{M}_D (H, J) = \cup_{\tilde{y} \in \tilde{P}(H)} \mathcal{M}_D (\tilde{y}; H, J). \]
As in (4.3) and (4.4) it is easy to prove that the union at the right side is only a finite union. That is, there only exist finitely many \( \tilde{y}_1, \ldots, \tilde{y}_r \) in \( \tilde{P}(H) \) such that \( \mathcal{M}_D (\tilde{y}_i; H, J) \neq \emptyset \) for \( i = 1, \ldots, r \). Thus
\[ (4.10) \quad \mathcal{M}_D (H, J) = \cup_{i=1}^r \mathcal{M}_D (\tilde{y}_i; H, J). \]
In particular, (4.3) and (4.4) imply that
\[ (4.11) \quad \mathcal{M}_{B-A} (H, J) = \cup_{i=1}^s \mathcal{M}_{B-A} (\tilde{x}_i; (-A); H, J). \]
Note that the right sides of both (4.10) and (4.11) are all disjoint unions. In order to compactify \( \mathcal{M}_D (H, J) \) we introduce:
Definition 4.2. Given \( \tilde{y} \in \overline{P}(H) \), \( D \in \Gamma \), a semistable \( \mathcal{F} \)-curve \((\Sigma, \ell)\) with at least two principal components (cf. Def. 2.1), a continuous map \( f : \Sigma \setminus \{z_1, \ldots, z_N\} \to M \) is called a stable \((J, H)\)-broken solution with a joint \( \tilde{y} \) and of class \( D \) if we may divide \((\Sigma, \ell)\) into two semistable \( \mathcal{F} \)-curve \((\Sigma_+, \ell^+)\) and \((\Sigma_-, \ell^-)\) at some double point \( z_i \) between two principal components, \( 2 \leq i_0 \leq N_p \), such that \( f|_{\Sigma_+} \) and \( f|_{\Sigma_-} \) are the stable \((J, H)\)+-disk with cap \( \tilde{y} \) and stable \((J, H)\)⁻-disk with cap \( \tilde{y}^2(\bar{z}) \) respectively. Furthermore, a tuple \((f, \Sigma, \ell)\) is called a stable \((J, H)\)-broken solution of class \( D \) if it is stable \((J, H)\)-broken solution with a joint \( \tilde{y} \) and of class \( D \) for some \( \tilde{y} \in \overline{P}(H) \).

It is easily seen that if a stable \((J, H)\)-broken solution \((f, \Sigma, \ell)\) with a joint \( \tilde{y} \) and of a class \( D \) has at least three principal components then it is also such a broken solution with another joint different from \( \tilde{y} \) and of class \( D \). As usual we may define the equivalence class of such a broken solution in an obvious way. But it should be noted that two equivalence classes of a given stable \((J, H)\)-broken solution \((f, \Sigma, \ell)\) with a joint \( \tilde{y} \) as a stable \((J, H)\)-broken solution with a joint \( \tilde{y} \) and as a stable \((J, H)\)-broken solution are same. We still denote by \((f)\) the equivalence class of \((f, \Sigma, \ell)\). Define the energy of such a \((f)\) by \( E_D((f)) = E_D(f) = E_+(f|_{\Sigma_+}) + E_-(f|_{\Sigma_-}) \). Then (2.6) yields that \( E_D((f)) \leq \omega(D) + 2 \max |H| \). Let us denote by \( \overline{M}_D(\tilde{y}; H, J) \) and \( \overline{M}_D(H, J) \) the spaces of all equivalence classes of the two kinds of maps respectively. Then

\begin{equation}
\overline{M}_D(H, J) = \bigcup_{\tilde{y} \in \overline{P}(H)} \overline{M}_D(\tilde{y}; H, J).
\end{equation}

However, one should also note that \( \overline{M}_D(\tilde{y}; H, J) \) and \( \overline{M}_D(\tilde{z}; H, J) \) have probably nonempty intersection for two different \( \tilde{y} \) and \( \tilde{z} \). Thus we cannot affirm the above unions to be the disjoint unions. But each stable \((J, H)\)-broken solution has at most \( 2n = \text{dim} M \) joints. As in (4.3) and (4.4) we can prove the union at the right side of (4.12) is actually a finite unions, i.e., other \( \overline{M}_D(\tilde{y}; H, J) \) are empty except finitely many \( \tilde{y} \in \overline{P}(H) \), saying \( \tilde{y}_1, \ldots, \tilde{y}_t \), \( t \geq r \) because of (4.10) and the fact that \( M_D(\tilde{y}; H, J) \subset \overline{M}_D(\tilde{y}; H, J) \). Then we get that

\begin{equation}
\overline{M}_D(H, J) = \bigcup_{i=1}^t \overline{M}_D(\tilde{y}_i; H, J).
\end{equation}

Note that by the above assumptions if \( t > r \) then \( M_D(\tilde{y}_i; H, J) = \emptyset \) for \( r < i \leq t \). Unlike in (4.10) the union at the right side is not necessarily disjoint. However, as in Proposition 2.5 one easily shows that the spaces \( \overline{M}_D(\tilde{y}; H, J) \) and \( \overline{M}_D(H, J) \) equipped with the weak \( C^\infty \)-topology are the Hausdorff compactifications of \( M_D(\tilde{y}; H, J) \) and \( M_D(H, J) \) respectively (cf. [LiuT1]). Moreover, the notions of the corresponding stable \( L^k \)-broken maps may also be introduced in the similar way. Let \( \overline{M}_D^0(\tilde{y}; H, J) \)
and $\mathcal{M}_D(H,J)$ be the corresponding virtual moduli cycles to them respectively. Consider the evaluation maps

$$EV_D^\nu(\bar{y}) : \mathcal{M}_D(\bar{y}; H, J) \to M \times M$$

and $EV_D^\nu : \mathcal{M}_D(H, J) \to M \times M$ given by $EV_D^\nu(\bar{y})(f) = (f(z_-), f(z_+))$ and $EV_D^\nu(f) = (f(z_-), f(z_+))$, where $z_-$ and $z_+$ two double points on the chain of the principal components of the domain of $f$ (cf. Def. 2.1). For the map $E_a^\nu \times E_b^\nu$ in (4.7), if

$$\mu(a) - \mu(b) + 2c_1(D) = 0,$$

then for a generic $(h, g) \in O(h_0) \times R$ we get two (rational) intersection numbers

$$n^\nu_D(a, b, \bar{y}) := (E_a^\nu \times E_b^\nu) \cdot EV_D^\nu(\bar{y})$$

and

$$n^\nu_D(a, b) := (E_a^\nu \times E_b^\nu) \cdot EV_D^\nu,$$

which are independent of choices of generic $\nu$ and $(h, g) \in O(h_0) \times R$.

**Proposition 4.3.** Under the above assumptions it holds that

$$n^\nu_D(a, b, \bar{y}) = n^{\nu^+}_D(a, \bar{y}) \cdot n^{\nu^-}_D(b, \bar{y}; (-D)),$$

$$n^\nu_D(a, b) = \sum_{i=1}^r n^\nu_D(a, b, \bar{y}_i).$$

From these it easily follows that (4.5) and (4.9) become

$$m^\nu_{+, -}(\langle a, A \rangle; \langle b, B \rangle) = (E_a^\nu \times E_b^\nu) \cdot EV_{B-A}^\nu.$$

**Proof of Proposition 4.3.** We first prove (4.16). By the definition in (3.2) the rational numbers $n^{\nu^+}_D(a, \bar{y})$ and $n^{\nu^-}_D(b, \bar{y}; (-D))$ are independent of the choices of generic small $\nu^+$ and $\nu^-$. Therefore, we only need to prove (4.16) for suitable regular $\nu$ and $\nu^\pm$. Notice that $\mathcal{M}_D(\bar{y}; H, J)$ can be identified with the product

$$\mathcal{M}_+(\bar{y}; H, J) \times \mathcal{M}_-(\bar{y}; (-D); H, J)$$

by the map $(f, \Sigma, \bar{y}) \mapsto (\langle f|_{\Sigma_+}, \Sigma_+, \bar{y}^+\rangle, \langle f|_{\Sigma_-}, \Sigma_-, \bar{y}^-\rangle)$. More generally, let

$$\mathcal{B}_D^{p,k}(\bar{y}; H), \quad \mathcal{B}_+^{p,k}(\bar{y}; H) \quad \text{and} \quad \mathcal{B}_-^{p,k}(\bar{y}; (-D); H)$$

be the corresponding $L^p_k$-stable map spaces then the first one can be identified with the product space $\mathcal{B}_+^{p,k}(\bar{y}; H) \times \mathcal{B}_-^{p,k}(\bar{y}; (-D); H)$ by the natural map

$$GL : (\langle f_+, \Sigma_+, \bar{y}^+\rangle, \langle f_-, \Sigma_-, \bar{y}^-\rangle) \mapsto \langle f_+ \cup_y f_-, \Sigma_+ \cup \Sigma_-, \bar{y}^+ \cup \bar{y}^-\rangle.$$
We shall prove that a virtual moduli cycle $\overline{\mathcal{M}}^\nu_D(\gamma; H, J)$ for $\overline{\mathcal{M}}_D(\gamma; H, J)$ naturally induces the virtual moduli cycle $\overline{\mathcal{M}}^\nu_+(\gamma; H, J)$ for $\overline{\mathcal{M}}_+(\gamma; H, J)$ and that $\overline{\mathcal{M}}^\nu_-(\gamma^\sharp(-D); H, J)$ for $\overline{\mathcal{M}}_-(\gamma^\sharp(-D); H, J)$ such that

\begin{equation}
\overline{\mathcal{M}}^\nu_D(\gamma; H, J)
= \{ f_+ g_f - (f_+, f_-) \in \overline{\mathcal{M}}^\nu_+(\gamma; H, J) \times \overline{\mathcal{M}}^\nu_-(\gamma^\sharp(-D); H, J) \}.
\end{equation}

Once it is proved. Note that for $f = f_+ g_f - (f_+, f_-) \in \overline{\mathcal{M}}^\nu_D(\gamma; H, J)$,

$EV^\nu_D(\gamma)(f) = (EV^\nu_+(\gamma)(f_+), EV^\nu_-(\gamma^\sharp(-D))(f_-))$.

By (4.15) and (4.20), for a generic choice of $(h, g) \in \mathcal{O}(h_0) \times \mathcal{R}$ we have

$$n^\nu_D(a, b, \tilde{y}) = (E^a \times E^b) : EV^\nu_D(\gamma)$$
$$= (E^a \times E^b) \cdot (EV^\nu_+(\gamma) \times EV^\nu_-(\gamma^\sharp(-D)))$$
$$= (E^a \cdot EV^\nu_+(\gamma)) \times (E^b \cdot EV^\nu_-(\gamma^\sharp(-D)))$$
$$= n^\nu_+(a, \tilde{y}) \cdot n^\nu_-(b, \gamma^\sharp(-D)).$$

Namely, (4.16) holds. In order to prove (4.20) we follow [LiuT1, LiuT2] and [LiuT3] to choose finitely many points $\langle f^+_i \rangle = \langle f^+_i, \Sigma^+_i, \tilde{f}^+_i \rangle \in \overline{\mathcal{M}}_+(\gamma; H, J)$ and $\langle f^-_j \rangle = \langle f^-_j, \Sigma^-_j, \tilde{f}^-_j \rangle \in \overline{\mathcal{M}}_-(\gamma^\sharp(-D); H, J)$, and their open neighborhoods $W^+_i = W^+_i(\langle f^+_i \rangle)$ in $B^k_+(\gamma; H)$ and those $W^-_j = W^-_j(\langle f^-_j \rangle)$ in $B^k_-(\gamma^\sharp(-D); H)$, $i = 1, \ldots, m_+$, $j = 1, \ldots, m_-$, such that the following hold:

(i) $W^+(\gamma) := \cup_{i=1}^{m_+} W^+_i$ and $W^-(\gamma^\sharp(-D)) := \cup_{j=1}^{m_-} W^-_j$ are, respectively, open neighborhoods of $\overline{\mathcal{M}}_+(\gamma; H, J)$ in $B^k_+(\gamma; H)$ and those of $\overline{\mathcal{M}}_-(\gamma^\sharp(-D); H, J)$ in $B^k_-(\gamma^\sharp(-D); H)$.

(ii) $W^+_i$ and $W^-_j$ have the uniformizers $(\tilde{W}^+_i, \Gamma^+_i, \pi^+_i)$ and $(\tilde{W}^-_j, \Gamma^-_j, \pi^-_j)$ respectively, $i = 1, \ldots, m_+$, $j = 1, \ldots, m_-$.

(iii) $\{ (\tilde{W}^+_i, \Gamma^+_i, \pi^+_i) \}_{i=1}^{m_+}$ and $\{ (\tilde{W}^-_j, \Gamma^-_j, \pi^-_j) \}_{j=1}^{m_-}$ constitute the orbifold atlases on $W^+(\gamma)$ and $W^-(\gamma^\sharp(-D))$ respectively.

(iv) There exist the local orbifold bundles $\tilde{\mathcal{L}}^+_i \to W^+_i$ and $\tilde{\mathcal{L}}^-_j \to W^-_j$ with uniformizing groups $\Gamma^+_i$ and $\Gamma^-_j$ respectively.

(v) There exist two obvious smooth sections

$$\tilde{\varphi}^+_{i, H} : \mathcal{W}^+(\gamma) \to \mathcal{L}^+_i = \cup_{i=1}^{m_+} \mathcal{L}^+_i$$

and

$$\tilde{\varphi}^-_{j, H} : \mathcal{W}^-(\gamma^\sharp(-D)) \to \mathcal{L}^-_j = \cup_{j=1}^{m_-} \mathcal{L}^-_j$$

such that their zero sets are $\overline{\mathcal{M}}_+(\gamma; H, J)$ and $\overline{\mathcal{M}}_-(\gamma^\sharp(-D); H, J)$ respectively, and that they may be lifted to the collections $\{ \tilde{\varphi}^+_i \}_{i=1}^{m_+}$ and $\{ \tilde{\varphi}^-_j \}_{j=1}^{m_-}$.
of $\Gamma^+_i$-equivariant sections of $\tilde{L}^+_i \to \tilde{W}^+_i$ and those $\{\tilde{\sigma}^{-j}_{J,H}\}_{j=1}^m$ of $\Gamma^-_j$-equivariant sections of $\tilde{L}^-_j \to \tilde{W}^-_j$ respectively.

As defined above (2.13) we have corresponding index sets

$$\mathcal{N}^\pm := \{I \subset \{1, \ldots, m^\pm\} | W^\pm_I = \cap_{k \in I} W^\pm_k \neq \emptyset\}.$$ 

For $I = \{i_1, \ldots, i_p\} \in \mathcal{N}^+$ and $I = \{j_1, \ldots, j_q\} \in \mathcal{N}^-$ we denote by $\Gamma^+_I = \prod_{k=1}^{p} \Gamma^+_{i_k}$ and $\Gamma^-_J = \prod_{l=1}^{q} \Gamma^-_{j_l}$. Correspondingly, we have also the fibre products

$$\tilde{W}_{I}^{I^+} = \{x_i^+ = (x_i^+)_{i \in I} | \pi_i^+(x_i^+) = \pi_{i'}(x_i^+) \in W^+_I, \forall i, i' \in I\},$$

$$\tilde{W}_{J}^{J^-} = \{y_j^- = (y_j^-)_{j \in J} | \pi_j^-(y_j^-) = \pi_{j'}(y_j^-) \in W^-_J, \forall j, j' \in J\}$$

and the projections $\pi^+_I : \tilde{W}_{I}^{I^+} \to W^+_I$ and $\pi^-_J : \tilde{W}_{J}^{J^-} \to W^-_J$. For $I, I' \in \mathcal{N}^+$ with $I \subset I'$ and $J, J' \in \mathcal{N}^-$ with $J \subset J'$ we have also obvious projections $\pi^+_I$ and $\pi^-_J$ satisfying $\pi^+_I \circ \pi^+_I = \pi^+_I$ and $\pi^-_J \circ \pi^-_J = \pi^-_J$. Following [LiuT3] we may construct the desingularizations of $\tilde{W}^{I^+}$ and $\tilde{W}^{J^-}$ as follows:

$$\tilde{W}^{I^+} = \{\tilde{W}_{I}^{I^+}, \pi^+_I | I \subset I' \in \mathcal{N}^+\}$$

$$\tilde{W}^{J^-} = \{\tilde{W}_{J}^{J^-}, \pi^-_J | J \subset J' \in \mathcal{N}^-\}.$$ 

By (2.15) and (2.16), for every $I \subset \mathcal{N}^+$ (resp. $J \subset \mathcal{N}^-$) there exists a subset $V^+_I \subset W^+_I$ (resp. $V^-_J \subset W^-_J$) such that:

1. $\overline{M}_+(\bar{y}; H, J) \subset \cup_{I \in \mathcal{N}^+} V^+_I$ (resp. $\overline{M}_-(\bar{y}; (-D); H, J) \subset \cup_{J \in \mathcal{N}^-} V^-_J$);

2. $V^+_I \cap V^+_{I'} \neq \emptyset$ (resp. $V^-_J \cap V^-_{J'} \neq \emptyset$) only when $I \subset I'$ (resp. $J \subset J'$).

Then one can obtain from $\tilde{W}_{I}^{I^+}$ (resp. $\tilde{W}_{J}^{J^-}$) the desingularization $\tilde{V}_{I}^{I^+}$ (resp. $\tilde{V}_{J}^{J^-}$) of the restriction of $\tilde{W}_{I}^{I^+}$ (resp. $\tilde{W}_{J}^{J^-}$) to $V^+_I$ (resp. $V^-_J$). Let us denote by

$$\tilde{V}^{I^+} = \{\tilde{V}_{I}^{I^+}, \pi^+_I | I \subset I' \in \mathcal{N}^+\}$$

$$\tilde{V}^{J^-} = \{\tilde{V}_{J}^{J^-}, \pi^-_J | J \subset J' \in \mathcal{N}^-\}.$$
With the same way we may use \( \tilde{L}^+ \) and \( \tilde{L}^- \) to construct \( \tilde{L}_{ij}^+ \), \( \tilde{L}_{ij}^- \) and \( \tilde{L}_{ij}^\Gamma \) for \( I \in \mathcal{N}^+ \) and \( J \in \mathcal{N}^- \), and obtain the following systems of bundles

\[
\begin{align*}
(\tilde{L}^+, \tilde{W}^+) &= \left\{ (\tilde{L}_{ij}^+, \tilde{W}_{ij}^+) \mid I \in \mathcal{N}^+ \right\} \\
(\tilde{L}^+, \tilde{V}^+) &= \left\{ (\tilde{L}_{ij}^+, \tilde{V}_{ij}^+) \mid I \in \mathcal{N}^+ \right\} , \\
(\tilde{L}^-, \tilde{W}^-) &= \left\{ (\tilde{L}_{ij}^-, \tilde{W}_{ij}^-) \mid J \in \mathcal{N}^- \right\} \quad \text{and} \\
(\tilde{L}^+, \tilde{V}^-) &= \left\{ (\tilde{L}_{ij}^+, \tilde{V}_{ij}^-) \mid J \in \mathcal{N}^- \right\} .
\end{align*}
\]

In particular, \( (\tilde{L}^+, \tilde{V}^+) \) and \( (\tilde{L}^-, \tilde{V}^-) \) are also the systems of the stratified Banach bundles. For \( i = 1, \ldots, m_+ \), \( j = 1, \ldots, m_- \), let \( \langle f_{ij} \rangle = \langle f_i^+ f_j^- \rangle \) and

\[
\begin{align*}
W_{ij} &= \{ (g^+ z_y g^-) \mid \langle g^+ \rangle \in W_i^+, \langle g^- \rangle \in W_j^- \} \\
\tilde{W}_{ij} &= \{ g^+ z_y g^- \mid g^+ \in \tilde{W}_i^+, g^- \in \tilde{W}_j^- \}.
\end{align*}
\]

Then \( \langle f_{ij} \rangle \in \mathcal{M}_D(\tilde{y}; H, J) \) and \( \{ W_{ij} \mid 1 \leq i \leq m_+, 1 \leq j \leq m_- \} \) is a covering of \( \mathcal{M}_D(\tilde{y}; H, J) \). Moreover, each \( \widetilde{W}_{ij} \) has a uniformizer \( (\tilde{W}_{ij}, \Gamma_{ij}, \pi_{ij}) \) with

\[
\Gamma_{ij} = \Gamma_i^+ \times \Gamma_j^- \quad \text{and} \quad \pi_{ij}(g^+ z_y g^-) = \text{GL}(\pi_i^+(g^+), \pi_j^-(g^-)).
\]

The action of \( \Gamma_{ij} \) on \( \tilde{W}_{ij} \) is given by \( \langle \phi^+, \phi^- \rangle \cdot \langle g^+, g^- \rangle = \langle \phi^+ \cdot g^+, \phi^- \cdot g^- \rangle \). As in §2 we may construct the local orbifold bundle \( \mathcal{L}_{ij} \rightarrow W_{ij} \) uniformized by \( \tilde{L}_{ij} \rightarrow \tilde{W}_{ij} \) with uniformizing group \( \Gamma_{ij} \). Moreover, there exists a natural section \( \partial_{D,H}^{ij} : W \rightarrow \mathcal{L} = \bigcup_{i,j} W_{ij} \) whose zero set is \( \mathcal{M}_D(\tilde{y}; H, J) \) and that lifts to a collection \( \{ \partial_{D,H}^{ij} \mid 1 \leq i \leq m_+, 1 \leq j \leq m_- \} \) of \( \Gamma_{ij} \)-equivariant sections of \( \tilde{L}_{ij} \rightarrow \tilde{W}_{ij} \). Here \( \partial_{D,H}^{ij}, \partial_{D,H}^{+,ij} \) and \( \partial_{D,H}^{-,ij} \) satisfy: For \( g = g^+ z_y g^- \in \tilde{W}_{ij} \) and \( \xi \in (\tilde{L}_{ij})_g = L^p_{k-1}(\wedge^{0,1}(g^* TM)) \) it must hold that

\[
\tilde{\partial}_{D,H}^{+,ij} \xi = (\partial_{D,H}^{+,ij}(\xi|_{\Sigma^+}))^\sharp (\partial_{D,H}^{-,ij}(\xi|_{\Sigma^-}))
\]

because two terms at the right side are the same when restricted to \( \Sigma^+ \cap \Sigma^- \). Here \( \Sigma^+ \) and \( \Sigma^- \) are the domains of \( g^+ \) and \( g^- \) respectively. For given \( f_+ \in \mathcal{B}_+^{0,k}(\tilde{y}; H) \) and \( f_- \in \mathcal{B}_-^{0,k}(\tilde{y}; H) \) the elements \( \xi^+ \in L^p_{k-1}(\wedge^{0,1}(f_+^* TM)) \) and \( \xi^- \in L^p_{k-1}(\wedge^{0,1}(f_-^* TM)) \) might not be glued into one of

\[
L^p_{k-1}(\wedge^{0,1}((f_+^* f_-)^* TM)).
\]

Thus we might not glue \( \mathcal{L}^+ \) and \( \mathcal{L}^- \) in general.
Also denote by $\mathcal{N} = \mathcal{N}^+ \times \mathcal{N}^-$. For $(I, J)$ and $(I', J')$ in $\mathcal{N}$ we say $(I, J) \subset (I', J')$ if $I \subset I'$ and $J \subset J'$. Corresponding to this covering we have $W_{(I,J)} = \text{GL}(W_I \times W_J)$, $\Gamma_{(I,J)} = \Gamma_I^+ \oplus \Gamma_J^-$ and the system of bundles
\[
(\mathcal{E}^+, \mathcal{V}^+) = \left\{ \left( \mathcal{E}^+_I, \mathcal{V}^+_I \right), \pi_{(I,J)}, \pi_{(I',J')} \mid (I, J) \subset (I', J') \in \mathcal{N} \right\},
\]
\[
(\mathcal{E}^-, \mathcal{V}^-) = \left\{ \left( \mathcal{E}^-_I, \mathcal{V}^-_I \right), \pi_{(I,J)}, \pi_{(I',J')} \mid (I, J) \subset (I', J') \in \mathcal{N} \right\}.
\]

Let $f_{ij} = f_i^+ \sharp g_j^-$ and $R(f_{ij})$ be the cokernel of $D \left( \mathfrak{D}_{j, H} \right) (f_{ij})$ in $(\mathcal{E}_{ij})_{f_{ij}}$ as before. Take smooth cut-off functions $\beta_\epsilon(f_i^+)$ on the domain $\Sigma_+^+$ of $f_i^+$ and $\beta_\epsilon(f_j^-)$ on the domain $\Sigma_-^-$ supported outside of the $\epsilon$-neighborhood of their double points. Then $\beta_\epsilon(f_i^+)$ and $\beta_\epsilon(f_j^-)$ naturally determine a smooth cut-off function, denoted by $\beta_\epsilon(f_{ij}) = \beta_\epsilon(f_i^+ \sharp \beta_\epsilon(f_j^-)$, on the domain $\Sigma_{ij} = \Sigma_+^+ \# \Sigma_-^-$ of $f_{ij}$ supported outside of the $\epsilon$-neighborhood of their double points. As in Step 1 in §2.2 and (2.19) we may use these to define corresponding spaces
\[
R_\epsilon(f_{ij}) \text{ and } R^\epsilon((f_{ij})) = \oplus_{i=1}^{\infty} \oplus_{j=1}^{\infty} R_\epsilon(f_{ij}).
\]

Now as in (2.18) we take the smooth $\Gamma_I^+$-invariant cut-off functions $\gamma(f_i^+)$ on $\mathcal{V}^+_i$ and $\gamma(f_j^-)$ on $\mathcal{V}^-_j$ such that for each $\nu_{ij} \in R_\epsilon(f_{ij})$ (which may be written as $\nu_i^+ \sharp \nu_j^-$ for some $\nu_i^+ \in R_\epsilon(f_i^+)$ with domain $\Sigma_i^+$ and $\nu_j^- \in R_\epsilon(f_j^-)$ with domain $\Sigma_j^-$),
\[
\gamma(f_i^+) \cdot \nu_{ij}|_{\Sigma_i^+} = \gamma(f_i^+) \cdot \nu_i^+ \quad \text{and} \quad \gamma(f_j^-) \cdot \nu_{ij}|_{\Sigma_j^-} = \gamma(f_j^-) \cdot \nu_j^-
\]
will give rise to the global section
\[
\nu_i^+ = \nu_{ij}|_{\Sigma_i^+} = \left\{ \left( \nu_{ij}|_{\Sigma_i^+} \right)_I = \left( \nu_i^+ \right)_I \mid I \in \mathcal{N}^+ \right\}
\]
of $(\mathcal{L}^+, \mathcal{V}^+)$ and that
\[
\nu_j^- = \nu_{ij}|_{\Sigma_j^-} = \left\{ \left( \nu_{ij}|_{\Sigma_j^-} \right)_J = \left( \nu_j^- \right)_J \mid J \in \mathcal{N}^- \right\}
\]
of $(\mathcal{L}^-, \mathcal{V}^-)$ respectively. Then $\gamma(f_i^+) \sharp \gamma(f_j^-) \cdot \nu_{ij}$ (defined by $\gamma(f_i^+) \sharp \gamma(f_j^-) \cdot \nu_{ij}(2g_i^- \sharp g_j^-) = \gamma(f_i^+) \sharp \gamma(f_j^-) \cdot \nu_{ij}(2g_i^- \sharp g_j^-)$ can yield a global section $\nu_{ij} = \{ (\nu_{ij})_{I,J} \mid (I, J) \in \mathcal{N} \}$ of $(\mathcal{L}, \mathcal{V})$. For $\delta > 0$ small enough, it follows as before that for a generic choice of $\nu \in R^\epsilon((f_{ij}))$ the global section
\[
S_{D,H}^\nu = \left\{ S_{D,I,J}^\nu = \mathfrak{D}^\nu_{I,J} + \nu_{(I,J)} \mid (I, J) \in \mathcal{N} \right\},
\]
transversal to the zero section. Fix such a regular $\nu$. We obtain a virtual moduli cycle
\[
\overline{\mathcal{M}}_D(\bar{y}; H, J) = \sum_{(I,J) \in \mathcal{N}} \frac{1}{\Gamma(I,J)} \left\{ \pi_{(I,J)} : \mathcal{M}_D^{\nu_{(I,J)}}(\bar{y}; H, J) \to \mathcal{W} \right\}.
\]
Here $\mathcal{M}_D^{\nu_{l,j}}(\bar{y}; H, J) = (S_D^{\nu_{l,j}})^{-1}(0)$. Note that $\nu$ can be expressed as 
$\nu_{ij} = \nu_j^+ \nu_j^- \in R_\epsilon(f_{ij}),$ $i = 1, \ldots, m_+$ and $j = 1, \ldots, m_-$. We put 

$$(4.21) \quad \mathfrak{m}^+ = \sum_{i=1}^{m_+} \nu_{ij}|_{\Sigma_i^+} = \sum_{i=1}^{m_+} \nu_i^+ \quad \text{and} \quad \mathfrak{m}^- = \sum_{j=1}^{m_-} \nu_{ij}|_{\Sigma_j^-} = \sum_{j=1}^{m_-} \nu_j^-.$$ 

They are the global section of $(\mathcal{L}^+, \mathcal{V}^+)$ and that of $(\mathcal{L}^-, \mathcal{V}^-)$ respectively. We assert that $\mathcal{D}_{J,H}^+ + \mathfrak{m}^+$ (resp. $\mathcal{D}_{J,H}^- + \mathfrak{m}^-$) is also transversal to the zero section of $(\mathcal{L}^+, \mathcal{V}^+)$ (resp. $(\mathcal{L}^-, \mathcal{V}^-)$). We only prove the assertion for $\mathcal{D}_{J,H}^+ + \mathfrak{m}^+$. Firstly, using (4.19) it is not hard to check that $f \in \mathcal{M}^{\nu_{l,j}}_D(\bar{y}; H, J)$ if and only if $f = f^+y_j f^-$ for some $f^+ \in (\mathcal{D}_{J,H}^+ + \mathfrak{m}^+)^{-1}(0)$ and $f^- \in (\mathcal{D}_{J,H}^- + \mathfrak{m}^-)^{-1}(0)$. Next, by (4.21) we have 

$$\mathfrak{m}^+ = \sum_{i=1}^{m_+} \nu_{ij}|_{\Sigma_i^+} = \left\{ \sum_{i=1}^{m_+} (\nu_{ij}|_{\Sigma_i^+})_I \mid I \in N^+ \right\} = \left\{ \sum_{i=1}^{m_+} (\nu_i^+)_I \mid I \in N^+ \right\}.$$ 

For a given $f^+ \in (\mathcal{D}_{J,H}^+ + \mathfrak{m}^+)^{-1}(0)$ we choose any $f^- \in (\mathcal{D}_{J,H}^- + \mathfrak{m}^-)^{-1}(0)$ and obtain a $f = f^+y_j f^- \in \mathcal{M}^{\nu_{l,j}}_D(\bar{y}; H, J)$. Note that we can always extend any $\xi^+ \in (\mathcal{L}^+_{J,H} + \mathfrak{m}^+)$ into an element $\xi \in (\mathcal{L}_{(l,j)} + \mathfrak{m}^+)$ since $\mathcal{D}_{J,H}^+ + \mathfrak{m}_{(l,j)} : \mathcal{L}_{(l,j)} \to \mathcal{L}_{(l,j)}$ is transversal to the zero section we have $\eta \in \mathcal{D}_{(l,j)}(\mathcal{V}_{(l,j)})$ such that $D(\mathcal{D}_{J,H}^+ + \mathfrak{m}_{(l,j)})(\eta) = \xi$. This implies that $D(\mathcal{D}_{J,H}^+ + \mathfrak{m}_{(l,j)})(\eta|_{\Sigma^+}) = \xi^+$. The assertion is proved. In particular we obtain two virtual moduli cycles 

$$\mathcal{M}^{\nu_{l,j}}_+(\bar{y}; H, J) = \sum_{I \in N^+} \frac{1}{|\Gamma_I^+|} \left\{ \pi_I^+ : \mathcal{M}^{\nu_{l,j}}_+(\bar{y}; H, J) \to \mathcal{W}^+ \right\},$$ 

$$\mathcal{M}^{\nu_{l,j}}_-(\bar{y}; (\bar{y}-D); H, J) = \sum_{J \in N^-} \frac{1}{|\Gamma_J^-|} \left\{ \pi_J^- : \mathcal{M}^{\nu_{l,j}}_-(\bar{y}; (\bar{y}-D); H, J) \to \mathcal{W}^- \right\}.$$ 

Here $\mathcal{M}^{\nu_{l,j}}_+(\bar{y}; H, J) = (\mathcal{D}_{J,H}^+ + \mathfrak{m}^+)^{-1}(0)$ and $\mathcal{M}^{\nu_{l,j}}_-(\bar{y}; (\bar{y}-D); H, J) = (\mathcal{D}_{J,H}^- + \mathfrak{m}^-)^{-1}(0)$. Now (4.19) and the facts that $\pi_{(l,j)} = \pi_I^+ \times \pi_J^-$ and $|\Gamma_I,J| = |\Gamma_I^+| \times |\Gamma_J^-|$ together lead to 

$$\mathcal{M}^{\nu_{l,j}}_D(\bar{y}; H, J) = \mathcal{M}^{\nu_{l,j}}_+(\bar{y}; H, J) \times \mathcal{M}^{\nu_{l,j}}_-(\bar{y}; (\bar{y}-D); H, J).$$ 

This is equivalent to (4.20).

Next we prove (4.17). The ideas are similar to the proof of (4.16). Denote by $\bar{y}_i = [y_i, w_i]$. By (4.13), for $i = 1, \ldots, t$, we take small neighborhoods $W_{(i)}^l$ centred at $(f_{(i)}^l) \in \mathcal{M}(\bar{y}_i; H, J)$ in $\mathcal{B}^{\nu_{l,j}}_D(\bar{y}_i; H, J)$, $l = l_{i-1} + 1, \ldots, l_{i-1} + l_i$ with $l_0 = 1$, such that $\{W_{(i)}^l\}_{l=l_{i-1}+1}^{l_i}$ constitutes a finite covering of $\mathcal{C}(\bar{y}_i; H, J)$ satisfying the requirements to construct the virtual moduli cycle. Let $L_{(i)}^{(l)} \to \mathcal{M}^{\nu_{l,j}}_D(\bar{y}_i; H, J)$.
$W_i^{(i)}$ be the local orbifold bundle uniformized by $\tilde{L}_i^{(i)} \to \tilde{W}_i^{(i)}$ with uniformizing group $\Gamma_i^{(i)}$ and projection $\pi_i^{(i)}$, and

$$\overline{\partial}_{I,H,D}^{(i)} : \mathcal{W}^{(i)} = \bigcup_{l=1}^{l_i} W^{(i)} \to \mathcal{L}^{(i)} = \bigcup_{l=1}^{l_i} \mathcal{L}_l^{(i)},$$

$$\overline{\partial}_{I,H}^{(i)} : \mathcal{W} = \bigcup_{i=1}^{t} \bigcup_{l=1}^{l_i} W^{(i)} \to \mathcal{L} = \bigcup_{i=1}^{t} \bigcup_{l=1}^{l_i} \mathcal{L}_l^{(i)}$$

be the obvious sections whose zero sets are $\overline{\mathcal{M}}_D(\hat{y}_i; H, J)$ and $\overline{\mathcal{M}}_D(\hat{y}_i; H, J)$ respectively. Let $\mathcal{N}^{(i)}$ be the nerve of the covering $\{W_i^{(i)}\}_{l=1}^{l_i+1}$ of $\overline{\mathcal{M}}_D(\hat{y}_i; H, J)$, $i = 1, \ldots, t$, and $\mathcal{N}$ be that of the covering $\bigcup_{i=1}^{t} \{W_i^{(i)}\}_{l=1}^{l_i+1}$ of $\overline{\mathcal{M}}_D(\hat{y}_i; H, J)$ as before. The elements of $\mathcal{N}^{(i)}$ and those of $\mathcal{N}$ are denoted by $I^{(i)}$ and $I$ respectively. Correspondingly, we have the bundle systems

$$(\hat{L}_I^{(i)}, \hat{V}_I^{(i)}) = \left\{ \left( L_I^{(i)}, V_I^{(i)} \right) \mid I^{(i)} \in \mathcal{N}^{(i)} \right\},$$

$$(\hat{L}_I, \hat{V}_I) = \left\{ (\hat{L}_I, \hat{V}_I) \mid I \in \mathcal{N} \right\}.$$

As in (2.18) let $R^\nu(\{I^{(i)}\}) = \bigoplus_{i=1}^{t} \bigoplus_{l=1}^{l_i+1} R^\nu(I^{(i)})$ be the corresponding finite dimensional space that is used to construct the virtual moduli cycles from $(\hat{L}_I, \hat{V}_I)$ and $\overline{\partial}_{I,H}^{(i)}$. Then for a generic small $\nu \in R^\nu(\{I^{(i)}\})$ we obtain a global section $\nu : \hat{V}_I \to \hat{L}_I$ such that $S_D^\nu = \{S_D^\nu = \overline{\partial}_{I,H}^{(i)} + \nu \mid I \in \mathcal{N} \}$ is transversal to the zero section. As above we may prove that $\nu$ induces a global section $\nu^{(i)} : \hat{V}_I^{(i)} \to \hat{L}_I^{(i)}$ such that

$$S_D^{\nu^{(i)}} = \left\{ S_D^{\nu^{(i)}} = \overline{\partial}_{I,H,D}^{(i)} + \nu^{(i)} \mid I^{(i)} \in \mathcal{N}^{(i)} \right\}$$

is also transversal to the zero section for each $i$. Let

$$\overline{\mathcal{M}}_D(H, J) = \sum_{I \in \mathcal{N}} \frac{1}{|\Gamma_I|} \left\{ \pi_I : \mathcal{M}_D^{\nu^{(i)}}(H, J) \to \mathcal{W} \right\}$$

and

$$\overline{\mathcal{M}}_D^{\nu^{(i)}}(\hat{y}_i; H, J) = \sum_{I^{(i)} \in \mathcal{N}^{(i)}} \frac{1}{|\Gamma_I^{(i)}|} \left\{ \pi_I^{(i)} : \mathcal{M}_D^{\nu^{(i)}}(H, J) \to \mathcal{W}^{(i)} \right\},$$

$i = 1, \ldots, t$, be the corresponding virtual moduli cycles, where

$$\mathcal{M}_D^{\nu^{(i)}}(H, J) = (S_D^{\nu^{(i)}})^{-1}(0) \quad \text{and} \quad \mathcal{M}_D^{\nu^{(i)}}(\hat{y}_i; H, J) = (S_D^{\nu^{(i)}})^{-1}(0).$$
By (4.10) the top strata of $\mathcal{M}_D'(H,J)$ can only contain those of $\mathcal{M}_D^{(i)}(\bar{y}_i; H,J)$, $i = 1, \ldots, r$. Other top strata are all empty. Thus

$$T\mathcal{M}_D'(H,J) = \bigcup_{i=1}^r T\mathcal{M}_D^{(i)}(\bar{y}_i; H,J).$$

The union is also disjoint because $\bar{y}_i$, $i = 1, \ldots, r$, are different. Since (4.14) implies that the intersections in (4.15) may only occur in the top strata we arrive at

$$\mathcal{E}_D'(u) \cdot EV_D' = \sum_{i=1}^r (\mathcal{E}_a^u \times \mathcal{E}_b^s) \cdot EV_D^{(i)}(\bar{y}_i).$$

This completes the proof of (4.17). \hfill \Box

**Step 3.** Now we need to introduce a kind of deformation spaces for understanding the right side of (4.18). For $\rho \geq 0$ and $f : \mathbb{R} \times S^1 \to M$ we define $\partial_{J,H}\rho f$ by

$$(4.22) \quad \partial_{J,H}\rho f(s,t) = \partial_s f(s,t) + J(f)(\partial_t f - (\beta_+(s + \rho + 1) \cdot \beta_+(\rho + 1 - s))X_H(t,f)) = 0,$$

where $\beta_+$ is as in §2. For such a map $f$ we define the energy of it by

$$(4.23) \quad E_\rho(f) := \int_{-\infty}^{\infty} \int_0^1 |\partial_s f|^2 g_J ds dt.$$

By the removable singularity theorem, $E_\rho(f) < +\infty$ implies that $f$ can be extended into a smooth map from $\mathbb{C}P^1 \approx \{-\infty\} \cup \mathbb{R} \times S^1 \cup \{\infty\}$ to $M$ which is also $J$-holomorphic near 0 and $\infty$. For a given $D \in \Gamma$ we denote by $\mathcal{M}_D(H_\rho,J)$ the space of all maps satisfying (4.22), having finite energy and representing the class $D$. To compactify it we introduce:

**Definition 4.4.** Given $D \in \Gamma$ and a semistable $\mathcal{F}$-curve $(\Sigma,l)$ with a unique principal component (cf. Def. 2.1), a continuous map $f : \Sigma \to M$ is called a stable $(J,H_\rho)$-map of a class $D$ if it represents $D$ in the usual sense and also satisfies:

1. On the unique principal component $P$ with cylindrical coordinate $(s,t)$, $f^P = f|_{P - \{z_-, z_+\}}$ satisfies: $\bar{\partial}_{J,H_\rho}f^P = 0$ and $E_\rho(f^P) < +\infty$.
2. The restriction $f^B_i$ of $f$ to each bubble component $B_i$ is $J$-holomorphic, and the domain of each homologically trivial bubble component is stable.

We may also define the equivalence class of such a map. Denote by $\mathcal{M}_D(H_\rho,J)$ the space of all equivalence classes of such maps. The energy of $f$ is defined by

$$E_\rho(f) = \int_{-\infty}^{\infty} \int_0^1 |\partial_s f^P|^2 g_J ds dt + \sum_i \int_{B_i} (f^B_i)^* \omega.$$
From the arguments in [Sch3, 4.2] it follows that
\[(4.24)\quad E_\rho(f) \leq \omega(D) + 2\max |H|.
\]
As usual, for any \(\rho \geq 0\) we may construct the virtual moduli cycle \(\overline{\mathcal{M}}_{B-A}^\rho(H_\rho, J)\) of dimension \(2n + 2c_1(B - A)\) corresponding to the space \(\mathcal{M}_{B-A}(H_\rho, J)\). Consider the evaluation
\[E_{B-A}^\rho : \overline{\mathcal{M}}_{B-A}^\rho(H_\rho, J) \to M \times M, \quad f \mapsto (f(z_-), f(z_+)).\]
By (4.2), for a generic \((h, g) \in \mathcal{O}(h_0) \times \mathcal{R}\) we get a rational intersection number
\[(4.25)\quad n_{B-A}^\rho(a, b, h, g; H_\rho, J) := (F^a \times E_b^g) \cdot E_{B-A}^\rho.
\]
Define \((\Psi \circ \Phi)_\rho : QC_*(M, \omega; h, g; \mathbb{Q}) \to QC_*(M, \omega; h, g; \mathbb{Q})\) by
\[(\Psi \circ \Phi)_\rho((a, A)) = \sum_{\mu((b, B)) = \mu((a, A))} n_{B-A}^\rho(a, b, h, g; H_\rho, J) \cdot (b, B).
\]

**Proposition 4.5.** For every \(\rho \geq 0\), \((\Psi \circ \Phi)_\rho\) is a chain homomorphism.

**Proof.** Firstly, note that if \(n_{B-A}^\rho(a, b, h, g; H_\rho, J) \neq 0\) then \(\overline{\mathcal{M}}_{B-A}(H_\rho, J) \neq \emptyset\) and thus it follows from (4.24) that \(0 \leq \omega(B - A) + 2\max |H|\). Using the fact, as in Step 1 we can prove that \((\Psi \circ \Phi)_\rho\) indeed maps \(QC_*(M, \omega; h, g; \mathbb{Q})\) into itself.

Next, we show that \((\Psi \circ \Phi)_\rho\) commutes with the boundary operator \(\partial^Q\) in (1.6). It suffices to prove that \(\partial \circ (\Psi \circ \Phi)_\rho((a, A)) = (\Psi \circ \Phi)_\rho \circ \partial^Q((a, A))\) for each \((a, A) \in \operatorname{Crit}(h) \times \Gamma_k\). The direct computation shows that the left equals
\[
\sum_{\mu((c, B)) = k-1} \left[ \sum_{\mu(d) = \mu(c) + 1} n(d, c) n_{B-A}^\rho(a, d, h, g; H_\rho, J) \right] \cdot (c, B),
\]
and the right does
\[
\sum_{\mu((c, B)) = k-1} \left[ \sum_{\mu(b) = \mu(a) - 1} n(a, b) n_{B-A}^\rho(b, c, h, g; H_\rho, J) \right] \cdot (c, B).
\]
Therefore we only need to prove
\[
\sum_{\mu(d) = \mu(c) + 1} n(d, c) n_{B-A}^\rho(a, d, h, g; H_\rho, J) = \sum_{\mu(b) = \mu(a) - 1} n(a, b) n_{B-A}^\rho(b, c, h, g; H_\rho, J).
\]
This can be proved as Propositions 3.5 and 3.6. In fact, by
\[
\mu(a) - \mu(c) + 2c_1(B - A) = \mu((a, A)) - \mu((c, B)) = 1,
\]
for a generic \((h,g) \in \mathcal{O}(h_0) \times \mathcal{R}\) the fibre product
\[
(W^u(a,h,g) \times W^s(c,h,g)) \times E^\rho_a \times E^\nu_c = E^{\rho \nu}_{B-A} M_{B-A}(H_\rho, J)
\]
is a collection of compatible local cornered smooth manifolds of dimension 1 and with the natural orientations. Its boundary is given by the union
\[
\bigcup_{\mu(b)=\mu(a)-1} n(a,b) \cdot (W^u(b,h,g) \times W^s(c,h,g)) \times R_{bc} \ M_{B-A}(H_\rho, J) \cup \bigcup_{\mu(d)=\mu(c)+1} n(d,c) \cdot (W^u(a,h,g) \times W^s(d,h,g)) \times R_{ad} \ M_{B-A}(H_\rho, J)
\]
where \(R_{bc}\) are representing \(\{0,1\}\)-dimensional boundary manifolds of \(M_{B-A}(H_\rho, J)\) and \(R_{ad}\) are representing \(E^\rho_a \times E^\nu_c = E^{\rho \nu}_{B-A}\). Hence as before the conclusions can follow from
\[
\kappa \partial((W^u(a,h,g) \times W^s(c,h,g)) \times E^\rho_a \times E^\nu_c = E^{\rho \nu}_{B-A} M_{B-A}(H_\rho, J)) = 0.
\]

\[\square\]

Our purpose is to prove that \((\Psi \circ \Phi)_0\) is chain homotopy equivalent to \(\Psi \circ \Phi\). To this goal let us consider the space
\[
M_{B-A}(\{H_\rho\}, J) = \bigcup_{\rho \geq 0} \{\rho\} \times M_{B-A}(H_\rho, J).
\]
From the arguments in [HS1, S1, S2] and [Sch2] it is not hard to derive that for any sequence \(u^m \in M_{B-A}(H_\rho_m, J)\) with \(\rho_m \to +\infty\) there must exist a subsequence (still denoted by \(u^m\)), finitely many elements \(\overline{x}_1, \ldots, \overline{x}_k \in \overline{\mathcal{P}}(H)\), and \(u_0 \in M_+(\overline{x}_1; \theta, J), u_j \in M(\overline{x}_j, \overline{x}_{j+1}; H, J), j = 1, \ldots, k - 1, u_k \in M_-(\overline{x}_k; H, J)\), and sequences \(-\rho_m - 1 \equiv s^0_m < \cdots < s^k_m \equiv \rho_m + 1\), such that \(u^m(s + s^i_m, t)\) converge modulo bubbling to \(u_i(s, t)\) for \(i = 0, \ldots, k\) (see [S2] for the precise definition of this term). Moreover, if \(w^l_i, l = 1, \ldots, i\), are all bubbles attached to \(u_i\), then the connected sum of all \(u_i\), \(w^l_i\), \(l = 1, \ldots, i\), \(i = 0, \ldots, k\), represents the class \(B - A\). This convergence result shows that the stabilized space of \(M_{B-A}(\{H_\rho\}, J)\) is given by
\[
(4.26) \quad \overline{M}_{B-A}(\{H_\rho\}, J) = \bigcup_{\rho \in [0, +\infty]} \{\rho\} \times \overline{M}_{B-A}(H_\rho, J),
\]
where \(\overline{M}_{B-A}(H_{+\infty}, J)\) is understood as \(\overline{M}_{B-A}(H, J)\) in (4.11) with \(D\) replaced by \(B - A\), and \([0, +\infty] = [0, +\infty] \cup \{+\infty\}\) is the compactification of \([0, +\infty]\) equipped with the structure of a bounded manifold obtained by requiring that
\[
h : [0, +\infty] \to [0, 1], t \mapsto t/\sqrt{1 + t^2}
\]
is a diffeomorphism. Then by the standard arguments we can prove that the space in (4.26) is compact and Hausdorff with respect to the weak \(C^\infty\)-topology. In addition, one has the obvious continuous evaluation map
\[
(4.27) \quad \Xi_{B-A} : \overline{M}_{B-A}(\{H_\rho\}, J) \to M \times M
\]
given by \(\Xi_{B-A}(\rho, \langle f \rangle) = \langle f(z_-), f(z_+)\rangle\) for \(\langle f \rangle \in \overline{M}_{B-A}(H_\rho, J)\) with \(\rho \in [0, +\infty]\), and \(\Xi_{B-A}(+\infty, \langle f \rangle) = \langle f(z_-), f(z_+)\rangle\) for \(\langle f \rangle \in \overline{M}_{B-A}(H_{+\infty}, J)\).
As above we can construct an associated virtual moduli cycle \( \overline{M}_{\mathcal{B}}^\nu(\{H_\rho\}, J) \) of dimension \( 2n + 2c_1(B - A) + 1 \) with the boundary
\[
\partial \overline{M}_{\mathcal{B}}^\nu(\{H_\rho\}, J) = (-\overline{M}_{\mathcal{B}}^{\nu_0}(H_0, J)) \cup \overline{M}_{\mathcal{B}}^{\nu_\infty}(H_{+\infty}, J).
\]
Moreover, the evaluation map in (4.27) can naturally be extended onto the virtual moduli cycle, denoted by
\[
\Xi_{\mathcal{B}}^\nu : \overline{M}_{\mathcal{B}}^\nu(\{H_\rho\}, J) \to M \times M.
\]
Let \( n^{\nu_0}_{\mathcal{B}}(a, b, h, g; H_0, J) \) (resp. \( n^{\nu_\infty}_{\mathcal{B}}(a, b, h, g; H_{+\infty}, J) \)) denote the intersection number of \( E_a^\nu \times E_b^\nu \) and the restriction of \( \Xi_{\mathcal{B}}^\nu \) to \( \overline{M}_{\mathcal{B}}^{\nu_0}(H_0, J) \) (resp. \( \overline{M}_{\mathcal{B}}^{\nu_\infty}(H_{+\infty}, J) \)). Note that \( \overline{M}_{\mathcal{B}}^{\nu_\infty}(H_{+\infty}, J) \) is just a virtual moduli cycle associated with the space \( \overline{M}_{\mathcal{B}}(H, J) \) defined in (4.12). By (4.18) we get
\[
\Psi \circ \Phi(\langle a, A \rangle) = \sum_{\mu((b, B)) = \mu((a, A))} n^{\nu_\infty}_{\mathcal{B}}(a, b, h, g; H_{+\infty}, J) \cdot \langle b, B \rangle.
\]
As in [F, SZ] and [S2] we wish to define a homomorphism
\[
\varphi : QC_*(M, \omega; h, g; \mathbb{Q}) \to QC_*(M, \omega; h, g; \mathbb{Q})
\]
such that
\[
\Psi \circ \Phi - (\Psi \circ \Phi)_0 = \partial^Q \varphi + \varphi \partial^Q.
\]
For \( h \in \mathcal{O}(h_0) \), \( \langle a, A \rangle \in (\text{Crit}(h) \times \Gamma)_k \) and \( \langle d, D \rangle \in (\text{Crit}(h) \times \Gamma)_{k+1} \), the equality \( \mu(a) - \mu(d) + 2c_1(D - A) + 1 = 0 \) implies that for a generic \( (h, g) \in \mathcal{O}(h_0) \times \mathcal{R} \) the evaluations \( E_a^\nu \times E_d^\nu \) and \( \Xi_{\mathcal{B}}^\nu \) are intersecting transversally. So the rational intersection number
\[
n_{D-A}(a, d, h, g; \{H_\rho\}, J) := (E_a^\nu \times E_d^\nu) \cdot \Xi_{D-A}^\nu
\]
is well-defined. Then it is not difficult to check that \( \varphi \) defined by
\[
\varphi(\langle a, A \rangle) = \sum_{\mu((d, D)) = \mu((a, A))} n_{D-A}(a, d, h, g; \{H_\rho\}, J) \cdot \langle d, D \rangle
\]
is an endomorphism of \( QC_*(M, \omega; h, g; \mathbb{Q}) \) and satisfies (4.29). That is,
\[
\Psi \circ \Phi(\langle a, A \rangle) - (\Psi \circ \Phi)_0(\langle a, A \rangle) = \partial^Q \varphi(\langle a, A \rangle) + \varphi \partial^Q(\langle a, A \rangle)
\]
for each \( \langle a, A \rangle \in \text{Crit}(h) \times \Gamma \). In fact, by the direct computation it suffice to prove
\[
n^{\nu_\infty}_{\mathcal{B}}(a, b, h, g; H_{+\infty}, J) - n^{\nu_0}_{\mathcal{B}}(a, b, h, g; H_0, J)
\]
\[= \sum_{\mu(c) = \mu(b) + 1} n_{B-A}(a, c, h, g; \{H_\rho\}, J)n(c, b)
\]
\[= \sum_{\mu(d) = \mu(a) - 1} n(a, d)n_{B-A}(d, b, h, g; \{H_\rho\}, J)
\]
for each \((b, B) \in \text{Crit}(h) \times \Gamma\) with \(\mu((b, B)) = \mu((a, A))\). To prove it we take a generic \((h, g) \in \mathcal{O}(h_0) \times \mathcal{R}\) so that the evaluations \(E_a^u \times E_b^s, E_a^t \times E_c^s\) and \(E_d^u \times E_b^s\) are transversal to \(\Xi_{b, A}'\). By Lemmas 3.1 and 3.2, for a generic \((h, g)\) the fibre product

\[
(W^u(a, h, g) \times W^s(b, h, g)) \times \mathcal{M}_{B-A}^p((H_\rho), J)
\]

is a collection of compatible local cornered smooth manifolds of dimension 1 and with the natural orientations. Its boundary is given by

\[
\left((-W^u(a, h, g) \times W^s(b, h, g)) \times \mathcal{E}_{a,b}^s \times \mathcal{E}_{b}^{p0} \times \mathcal{M}_{B-A}^p((H_0, J)) \right)
\]

\[
\bigcup \bigcup \bigcup (W^u(a, h, g) \times W^s(b, h, g)) \times \mathcal{E}_{a,b}^s \times \mathcal{E}_{b}^{p0} \times \mathcal{M}_{B-A}^{p+\infty}(H_{+\infty}, J))
\]

\[
\bigcup \bigcup \bigcup n(a,d) : (W^u(a, h, g) \times W^s(d, h, g)) \times R_ad \times \mathcal{M}_{B-A}^p((H_\rho), J))
\]

\[
\bigcup \bigcup \bigcup n(c,b) : (-W^u(c, h, g) \times W^s(b, h, g)) \times R_cb \times \mathcal{M}_{B-A}^p((H_\rho), J))
\]

Here \(R_ad = \mathcal{E}_{a,b}^u \times \mathcal{E}_{b}^s = \Xi_{b, A}'\) and \(R_cb = \mathcal{E}_{a,b}^u \times \mathcal{E}_{b}^s = \Xi_{b, A}'\). This implies (4.31). To sum up we have proved:

**Proposition 4.6.** \(\Psi \circ \Phi\) is chain homotopy equivalent to \((\Psi \circ \Phi)_0\).

**Step 4.** We need to make further homotopy. For \(\tau \in [0, 1]\) we define

\[
\bar{\partial}_{J, \tau H_0} u(s, t) = \partial_s u(s, t) + J(u)(\partial_t u - \tau(\beta_+ (s + 1) - \beta_+(1 - s)) X_H(t, u)) = 0.
\]

In Definition 4.4 we replace \(\bar{\partial}_{J, H_0}\) with \(\bar{\partial}_{J, \tau H_0}\) and define the corresponding stable \((J, \tau H_0)\)-map of class \(D\). Let \(\mathcal{M}_D(\tau H_0, J)\) be the space of all equivalence classes of such maps. For \((f) \in \mathcal{M}_{B-A}(\tau H_0, J)\), as in (4.24) we can estimate

\[
E^\tau(f) := \int_0^1 |\partial_s f|^2_{g, J} ds dt + \sum_i \int_{B_i} (f B_i)^* \omega \leq \omega(B - A) + 2\tau \max |H|.
\]

As above we can construct a virtual moduli cycle \(\overline{\mathcal{M}}_{B-A}^\tau((\tau H_0), J)\) of the compact space \(\bigcup_{\tau \in [0, 1]} \{\tau\} \times \mathcal{M}_{B-A}(\tau H_0, J)\) of dimension \(2n + 2c_1(B - A) + 1\) and with boundary

\[
\partial \overline{\mathcal{M}}_{B-A}^\tau((\tau H_0), J) = (-\overline{\mathcal{M}}_{B-A}^\tau(0, J)) \cup \overline{\mathcal{M}}_{B-A}^\tau(H_0, J).
\]

(Actually, \(\overline{\mathcal{M}}_{B-A}^\tau(H_0, J)\) can be chosen as \(\overline{\mathcal{M}}_{B-A}^\tau(H_0, J)\).) For \(\tau = 0, 1\) and a generic \((h, g) \in \mathcal{O}(h_0) \times \mathcal{R}\), using the evaluation map

\[
E^\tau_{B-A} : \overline{\mathcal{M}}_{B-A}^\tau(\tau H_0, J) \to M \times M, f \mapsto (f(z_-), f(z_+))
\]

we get the well-defined rational intersection number

\[
n^\tau_{B-A}(a, b, h, g; \tau H_0, J) := (E^u_a \times E^s_b) \cdot E^\tau_{B-A}.
\]
By (4.25) we have
\[ n^\nu_1 B^{-A}(a, b, h, g; H_0, J) = n^\nu_0 B^{-A}(a, b, h, g; H_0, J) \]
since they are independent of the generic choices of \( \nu \) and \( (g, h) \). Therefore, as in Step 3 we can easily prove that \((\Psi \circ \Phi)_0\) and thus \(\Phi \circ \Phi\) are chain homotopy equivalent to \((\Psi \circ \Phi)^0\) defined by
\[
(\Psi \circ \Phi)^0(\langle a, A \rangle) = \sum_{\mu((b, B))=\mu((a, A))} n^\nu_0 B^{-A}(a, b, h, g; 0, J) \cdot \langle b, B \rangle.
\]
Here as in Proposition 4.5 it can be proved that \((\Psi \circ \Phi)_0\) is a chain homomorphism. We omit it. Now Theorem 4.1 can follow from this and the following result:

**Proposition 4.7.** The numbers \( n^\nu_0 B^{-A}(a, b, h, g; 0, J) \) at (4.32) satisfies
\[
n^\nu_0 B^{-A}(a, b, h, g; 0, J) = \begin{cases} 
1 & \text{if } a = b \text{ and } A = B, \\
0 & \text{otherwise}.
\end{cases}
\]

**Proof.** Case 1: \( A \neq B \). In this case \( \overline{M}_{B^{-A}}(0, J) \) contains no constant maps. Moreover, the domain of elements of \( \overline{M}_{B^{-A}}(0, J) \) is only 0-pointed semistable \( \mathcal{F} \)-curves with at least a principal component and the operator \( \overline{\partial}_J \) is invariant under action of the automorphism group of the domain of a semistable \( \mathcal{F} \)-curve. Thus as in [LiuT2] the associated virtual moduli cycle \( \overline{M}_{B^{-A}}^0(0, J) \) can be required to carry a free \( \mathbb{S}^1 \)-action under which the evaluation is invariant. Hence for a generic \( (h, g) \in \mathcal{O}(h_0) \times \mathcal{R} \) the fibre product
\[
(\overline{W}^u(a, h, g) \times \overline{W}^s(b, h, g)) \times_{(E_u \times E_s)=E_{B^{-A}}} \overline{M}_{B^{-A}}^0(0, J)
\]
must be empty under the condition (4.2). We get \( n^\nu_0 B^{-A}(a, b, h, g; 0, J) = 0 \).

Case 2: \( A = B \). Now \( \overline{M}_{B^{-A}}(0, J) \) can naturally be identified with \( M \) and thus \( \overline{M}_{B^{-A}}^0(0, J) \) may be taken as \( \overline{M}_{B^{-A}}(0, J) \). Note that \( \mu(a) = \mu(b) \) in the present case and that \( \overline{W}^u(a, h, g) \cap \overline{W}^s(b, h, g) \) carries a free \( \mathbb{R} \)-action if it is nonempty. The conclusions follow naturally. \( \square \)

Summing up the above arguments we complete the proof of Theorem 4.1.

**Remark 4.8.** If the Morse function \( h \) has only critical points of even index then \( \Psi \) is a right inverse of \( \Phi \) as the chain homomorphisms between \( QC_*(M, \omega; h, g; \mathbb{Q}) \) and \( C_*(H, J, \nu; \mathbb{Q}) \). Indeed, carefully checking the proof of Th. 4.1 one will find that \( n^{\nu^\infty}_B(a, b, h, g; H_+^\infty, J) - n^{\nu^0}_B(a, b, h, g; H_0, J) = 0 \) in (4.31). The same reasoning yields \( n^\nu_1 B^{-A}(a, b, h, g; H_0, J) = n^\nu_0 B^{-A}(a, b, h, g; 0, J) \). By Proposition 4.7 and (4.28) we get that \( \Psi \circ \Phi(\langle a, A \rangle) = \langle a, A \rangle \) for each \( \langle a, A \rangle \).
Theorem 4.9. \( \Phi \circ \Psi \) is chain homotopy equivalent to the identity. Consequently, \( \Phi \) induces a surjective \( \Lambda_{\omega} \)-module homomorphism from \( QH_*(h, g; \mathbb{Q}) \) to \( HF_*(M, \omega; H, J, \nu; \mathbb{Q}) \).

Proof. The proof is similar to that of Theorem 4.1. We only give main steps.

Step 1. For every \( \bar{x} \in \bar{P}_k(H) \) the direct computation gives

\[
\Phi \circ \Psi(\bar{x}) = \sum_{\mu(y)=k} m_{\nu,+}(\bar{x}, y) \cdot \bar{y},
\]

\[
m_{\nu,+}(\bar{x}, y) = \sum_{\mu(a,A)=k} n_{\nu,-}(a, \bar{x}^\nu(-A)) \cdot n_{\nu,+}(a, \bar{y}^\nu(-A)).
\]

As in the proof of Theorem 4.1 using Proposition 3.4 we can easily prove the second sum to be finite. That is, there are only finitely many \( a, A \in \Gamma \), saying \( A_1, \ldots, A_s \), such that

\[
\overline{M}_-(\bar{x}^\nu(-A_i); H, J) \neq \emptyset \quad \text{and} \quad \overline{M}_+(\bar{y}^\nu(-A_i); H, J) \neq \emptyset
\]

for \( i = 1, \ldots, s \). Note that the product \( n_{\nu,-}(a, \bar{x}^\nu(-A)) \cdot n_{\nu,+}(a, \bar{y}^\nu(-A)) \) can be explained as the intersection number of the product evaluations

\[
EV_{\nu,-} \times EV_{\nu,+} : \overline{M}_-(\bar{x}^\nu(-A); H, J) \times \overline{M}_+(\bar{y}^\nu(-A); H, J) \to M \times M
\]

and

\[
(E_a^s \times E_a^u \cdot (EV_{\nu,-}^{\nu,-}(\bar{x}^\nu(-A)) \times EV_{\nu,+}^{\nu,+}(\bar{y}^\nu(-A))))
\]

for a generic \( (h, g) \in \mathcal{O}(h_0) \times \mathcal{R} \), and that the fibre product

\[
(W^s(a, h, g) \times W^u(a, h, g)) \times_R (\overline{M}_-(\bar{x}^\nu(-A); H, J) \times \overline{M}_+(\bar{y}^\nu(-A); H, J))
\]

is an empty set for a generic \( (h, g) \in \mathcal{O}(h_0) \times \mathcal{R} \) even if \( \mu(\langle a, A \rangle) \neq k = \mu(\bar{x}) = \mu(\bar{y}) \). Here \( R = E_a^s \times E_a^u = EV_{\nu,-} \times EV_{\nu,+} \). Thus

\begin{equation}
(4.33)
\begin{cases}
m_{\nu,+}(\bar{x}, y) \\
= \sum_{(a,A) \in \text{Crit}(h) \times \Gamma} (E_a^s \times E_a^u) \cdot (EV_{\nu,-}^{\nu,-}(\bar{x}^\nu(-A)) \times EV_{\nu,+}^{\nu,+}(\bar{y}^\nu(-A))) \\
= \sum_{a \in \text{Crit}(h)} \sum_{i=1}^{s} (E_a^s \times E_a^u) \cdot (EV_{\nu,-}^{\nu,-}(\bar{x}^\nu(-A_i)) \times EV_{\nu,+}^{\nu,+}(\bar{y}^\nu(-A_i))).
\end{cases}
\end{equation}

Step 2. To understand this sum, for \( (h, g) \in \mathcal{O}(h_0) \times \mathcal{R} \) and \( \rho \geq 0 \) we denote by

\[
\mathcal{M}_\rho(h, g) := \{ \gamma \in C^\infty([-\rho, \rho], M) \mid \dot{\gamma} + \nabla_g h(\gamma) = 0 \}.
\]

It is a compact manifold of dimension \( 2n \) and \( \mathcal{M}_0(h, g) \) can naturally be identified with \( M \). Using the gluing techniques in the Morse homology (cf.
Theorem 6.8 in [Lu2]), the natural weak compactification of the noncompact manifold $\bigcup_{\rho \geq 0} \mathcal{M}_\rho(h, g)$ of dimension $2n + 1$ is given by

$$\bigcup_{\rho \geq 0} \mathcal{M}_\rho(h, g) := \bigcup_{\rho = 0}^\infty \mathcal{M}_\rho(h, g),$$

where

$$\mathcal{M}_\infty(h, g) := \bigcup_{a \in \text{Crit}(h)} \overline{\mathcal{W}}^s(a, h, g) \times \overline{\mathcal{W}}^u(a, h, g)$$

and the weak convergence of a sequence $\{\gamma_m\} \subset \mathcal{M}_{\rho_m}(h, g), \rho_m \to \infty$, towards a pair $(u, v) \in \mathcal{M}_\infty(h, g)$ is understood in an obvious way (cf. [AuB], [Sch1] and [Sch4]). The space has the structure of a manifold with corners and with boundary

$$\partial \bigcup_{\rho \geq 0} \mathcal{M}_\rho(h, g) = (-\mathcal{M}_0(h, g)) \cup (\bigcup_{a \in \text{Crit}(h)} \overline{\mathcal{W}}^s(a, h, g) \times \overline{\mathcal{W}}^u(a, h, g)).$$

Moreover, there exists a smooth evaluation

$$\text{ev}_{h, g} : \bigcup_{\rho \geq 0} \mathcal{M}_\rho(h, g) \to M \times M$$

given by $\text{ev}_{h, g}(\gamma) = (\gamma(-\rho), \gamma(\rho))$ for $\gamma \in \mathcal{M}_\rho(h, g)$ such that

$$\text{ev}_{h, g}|_{\mathcal{W}^s(a, h, g) \times \mathcal{W}^u(a, h, g)} = E^s_a \times E^u_a \text{ for each } a \in \text{Crit}(h).$$

We also denote by $\text{ev}^0_{h, g}$ the restriction of $\text{ev}_{h, g}$ to $\mathcal{M}_\rho(h, g)$ for $0 \leq \rho \leq \infty$. To make further arguments we need to assume that

$$\text{(4.34)} \quad \text{EV}_-^\nu(\tilde{x}^\sharp(-A_i)) \cap \text{EV}_+^\nu(\tilde{y}^\sharp(-A_i)), \quad i = 1, \ldots, s.$$  

These can actually be obtained for a generic small $(\nu^-, \nu^+) \in R^- \times R^+$ by increasing some points $f^-_j \in \overline{\mathcal{M}}_-(\tilde{x}^\sharp(\nu^+; H, J))$ and $f^+_j \in \overline{\mathcal{M}}_+(\tilde{y}^\sharp(\nu^-; H, J))$ and enlarging the spaces $R^\pm$ in the construction of the virtual moduli cycles. Using these, for a generic pair $(h, g) \in \mathcal{O}(h_0) \times \mathcal{R}$ the fibre product

$$\bigcup_{\rho \geq 0} \mathcal{M}_\rho(h, g) \times R_3 \left( \bigcup_{i=1}^n \overline{\mathcal{M}}_-^\nu(\tilde{x}^\sharp(-A_i); H, J) \times \overline{\mathcal{M}}_+^\nu(\tilde{y}, w^\sharp(-A_i); H, J) \right)$$

is a collection of compatible local cornered smooth manifolds of dimension 1 and with the natural orientations. Here $R_3$ represents $ev(h, g) = EV_-^\nu \times$
Proposition 4.10. \( \Phi \circ \Psi \) is chain homotopy equivalent to the homomorphism defined by

\[
\bar{\Phi} \circ \Psi(x) = \sum_{\mu(\bar{y})=\mu(x)} \left( \sum_{i=1}^{s} \text{EV}_-^{-}(\bar{x}^{-}_i(-A_i)) \cdot \text{EV}_+^{+}(\bar{y}^{+}_i(-A_i)) \right) \cdot \bar{y}.
\]
Now we also need to prove that $\Phi \circ \Psi$ is a chain homomorphism from $C_*(H, J; \mathbb{Q})$ to itself yet. The ideas are the same as those of Proposition 4.5. Let us outline its proof as follows: Firstly, (2.6) implies that if the sum in (4.35) is not zero then

$$\mathcal{F}_H(y) \leq \max |H| - \min_{1 \leq i \leq s} \omega(A_i) \quad \text{and} \quad \mathcal{F}_H(x) \geq -\max |H| - \max_{1 \leq i \leq s} \omega(A_i).$$

From these it easily follows that $\Phi \circ \Psi$ maps $C_*(H, J; \mathbb{Q})$ to itself. Next, by the direct computation we easily reduce the proof of $\Phi \circ \Psi \circ \partial F = \partial F \circ \Phi \circ \Psi$ to proving that for given $x \in \mathcal{P}_k(H)$ and $y \in \mathcal{P}_{k-1}(H)$ the following holds:

$$\sum_{i=1}^{s} \sum_{\mu(y)=k} (\text{EV}^-_i(x)(-A_i)) \cdot \text{EV}^+_i(y)(-A_i)) \cdot n_i(y, z) = \sum_{i=1}^{s} \sum_{\mu(z)=k-1} n_i(x, z') \cdot (\text{EV}^-_i(z')(A_i)) \cdot \text{EV}^+_i(z)(A_i)).$$

Here and in the following unions $n_i(x, z') = \sharp(C(\mathcal{M}^-_i(x)(-A_i), z')(A_i)))$ and $n_i(y, z) = \sharp(C(\mathcal{M}^+_i(y)(A_i), z)(-A_i))).$ In fact, by (4.34) the fibre product

$$\mathcal{M}^-_i(x)(A_i)) \times_{\text{EV}^-_i=\text{EV}^+_i} \mathcal{M}^+_i(z)(A_i))$$

is a collection of compatible local cornered smooth manifolds of dimension 1 and with the natural orientations. Its boundary is given by

$$\left( \bigcup_{\mu(y)=k-1} n_i(x, z') \cdot \mathcal{M}^-_i(z')(A_i)) \times_{\text{EV}^-_i=\text{EV}^+_i} \mathcal{M}^+_i(z)(A_i)) ight) \cup \left( - \bigcup_{\mu(y)=k} (\mathcal{M}^-_i(x)(A_i)) \times_{\text{EV}^-_i=\text{EV}^+_i} \mathcal{M}^+_i(y)(A_i)) \cdot n_i(y, z') \right).$$

Notice that the sum in (4.35) is exactly

$$\sharp \left( \bigcup_{i=1}^{s} \mathcal{M}^-_i(x)(A_i)) \times_{\text{EV}^-_i=\text{EV}^+_i} \mathcal{M}^+_i(y)(A_i)) \right).$$

They together lead to the desired conclusions.

**Step 3.** To understand the number in (4.35) we introduce:

**Definition 4.11.** Given $x, y \in \mathcal{P}(H)$ and a semistable $\mathcal{F}$-curve $(\Sigma, l)$ with at least two principal components (cf. Def. 2.1), a continuous map

$$f : \Sigma \setminus \{z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_{N_p+1}\} \to M$$

is called a **stable $(J, H)$-broken trajectory** if we divide $(\Sigma, l)$ into two semistable $\mathcal{F}$-curve $(\Sigma_-, l^-)$ and $(\Sigma_+, l^+)$ at some double point $z_i$ between two principal components, $1 < i < N_p + 1$, then $f|_{\Sigma_-}$ and $f|_{\Sigma_+}$ are the stable $(J, H)_-$-disk with cap $\tilde{x}$ and stable $(J, H)_+$-disk with cap $\tilde{y}$ respectively.
We can also define its equivalence class in an obvious way. Denote the space of all such equivalence classes by $\overline{\mathcal{M}}_{-+}(\vec{x}, \vec{y}; H, J)$. Let $\mathcal{M}_{-+}(\vec{x}, \vec{y}; H, J)$ be its subspace consisting of those elements whose domains have only two principal components and have no any bubble components. Then the former is the natural compactification of the latter, and has the virtual dimension $\mu(\vec{x}) - \mu(\vec{y})$. We can, as before, construct an associated virtual moduli cycle $\overline{\mathcal{M}}^\nu_{-+}(\vec{x}, \vec{y}; H, J)$ of dimension $\mu(\vec{x}) - \mu(\vec{y})$. Specially, if $\mu(\vec{x}) = \mu(\vec{y})$ this virtual moduli cycle determines a well-defined rational number $n^\nu_{-+}(\vec{x}, \vec{y}; H, J)$ which is independent of a generic choice of $\nu$ in the obvious way. As in the proof of Theorem 4.1, by carefully checking the construction of the virtual moduli cycle we can prove that

$$n^\nu_{-+}(\vec{x}, \vec{y}; H, J) = E^\nu_-(\vec{x}) \cdot E^\nu_+(\vec{y}).$$

Moreover, as showed in Step 1 there exist only finitely many $A_i \in \Gamma$ such that $\overline{\mathcal{M}}_{-+}(\vec{x}_i^\nu(-A_i), \vec{y}_i^\nu(-A_i); H, J) \neq \emptyset, i = 1, \ldots, s$. So the virtual moduli cycle associated with $\cup_{i=1}^s \overline{\mathcal{M}}_{-+}(\vec{x}_i^\nu(-A_i), \vec{y}_i^\nu(-A_i); H, J)$ can be taken as

$$\cup_{i=1}^s \overline{\mathcal{M}}^\nu_{-+}(\vec{x}_i^\nu(-A_i), \vec{y}_i^\nu(-A_i); H, J).$$

It follows that

$$\sharp(\cup_{i=1}^s \overline{\mathcal{M}}_{-+}(\vec{x}_i^\nu(-A_i), \vec{y}_i^\nu(-A_i); H, J))^\nu = \sum_{i=1}^s n^\nu_{-+}(\vec{x}_i^\nu(-A_i), \vec{y}_i^\nu(-A_i); H, J).$$

Thus (4.36) becomes

$$(4.37) \quad \Phi \circ \Psi(\vec{x}) = \sum_{\mu(\vec{y})=\mu(\vec{x})} \sharp(\cup_{i=1}^s \overline{\mathcal{M}}_{-+}(\vec{x}_i^\nu(-A_i), \vec{y}_i^\nu(-A_i); H, J))^\nu \cdot \vec{y}.$$

Step 4. Furthermore, for $\rho \geq 0$ and $f : \mathbb{R} \times S^1 \to M$ we define $\partial_{J_H^\rho} f$ by

$$\partial_{J_H^\rho} f(s,t) = \partial_s f(s,t) + J(f) (\partial_t f - (\beta_+(s-\rho) + \beta_+(s+\rho)) X_H(t,f)) = 0,$$

where $\beta_+$ is as in (4.22). For such a map $f$ we still define the energy of it by (4.23).

Definition 4.12. Given $[x, v], [y, w] \in \overline{\mathcal{P}}(H)$ and a semistable $\mathcal{F}$-curve $(\Sigma, \ell)$ as in Definition 2.1, a continuous map $f : \Sigma \setminus \{z_1, \ldots, z_{N_+} \} \to M$ is called a stable $(J, H)^\rho$-trajectory if there exist $[x, v] = [x_1, u_1], \ldots, [x_{N_+}, u_{N_+}] = [y, w]$ such that (1), (2), (3) in Def. 2.1 are satisfied unless (i) in Def. 2.1(1) is replaced by $\partial_{J_H^\rho} f_k = 0$ and $E_\rho(f_k^F) < +\infty$ in some principal component $P_k$.

Let $\overline{\mathcal{M}}(\vec{x}, \vec{y}; J, H^\rho)$ denote the space of equivalence classes of all stable $(J, H)^\rho$-trajectories from $\vec{x}$ to $\vec{y}$. As in §2 we can prove that this space is compact according to the weak $C^\infty$-topology and construct the corresponding virtual moduli cycle $\overline{\mathcal{M}}(\vec{x}, \vec{y}; J, H^\rho)$ of dimension $\mu(\vec{x}) - \mu(\vec{y})$. Specially, if $\mu(\vec{x}) = \mu(\vec{y})$ we can associate a rational number to it, denoted

$$n^\nu_{-+}(\vec{x}, \vec{y}; H, J) = E^\nu_-(\vec{x}) \cdot E^\nu_+(\vec{y}).$$
by $m'(\tilde{x}, \tilde{y}; J, H^p)$. Consider the space $\cup_{\rho \geq 0} \mathcal{M}(\tilde{x}, \tilde{y}; J, H^p)$. As before one easily shows that the natural weak compactification of it is given by

$$\cup_{\rho \geq 0} \mathcal{M}(\tilde{x}, \tilde{y}; J, H^p) \cup \bigcup_{i=1}^{s} \mathcal{M}_{-\epsilon}(\tilde{x}^{\#}(-A_i), \tilde{y}^{\#}(-A_i); H, J).$$

Using it we can construct a virtual moduli cycle of dimension 1 and prove:

**Proposition 4.13.** $\Phi \circ \Psi$ in (4.37) is chain homotopy equivalent to the homomorphism given by

$$(\Phi \circ \Psi)^0(\tilde{x}) = \sum_{\mu(\tilde{y}) = \mu(\tilde{x})} m'(\tilde{x}, \tilde{y}; J, H^0) \cdot \tilde{y}.$$

Here we have assumed that $(\Phi \circ \Psi)^0$ is a chain homomorphism from $C_*(H, J; \mathbb{Q})$ to itself. It can be proved as above. We omit it. As in [F, SZ] and [LiuT1], by taking a regular homotopy from $(J, H)$ to $(J, (\beta_+(\cdot) + \beta_-(\cdot))H)$ we can prove that $(\Phi \circ \Psi)^0$ is chain homotopy equivalent to the homomorphism defined by

$$(\Phi \circ \Psi)^0(\tilde{x}) = \sum_{\mu(\tilde{y}) = \mu(\tilde{x})} \sharp(C(\mathcal{M}^{\nu}(\tilde{x}, \tilde{y}; J, H))) \cdot \tilde{y}.$$

Since $\mu(\tilde{y}) = \mu(\tilde{x})$ it is easily checked that $\sharp(C(\mathcal{M}^{\nu}(\tilde{x}, \tilde{y}; J, H))) = 1$ as $\tilde{x} = \tilde{y}$, and $\sharp(C(\mathcal{M}^{\nu}(\tilde{x}, \tilde{y}; J, H))) = 0$ otherwise. That is, $(\Phi \circ \Psi)^0 = id$.

This fact, Proposition 4.13, (4.37) and Proposition 4.10 together prove Theorem 4.9.

**Acknowledgements.** The author is very grateful to Professors Gang Tian and Dietmar Salamon for their help in his understanding of Floer homology in past years. He would also like to thank Professors Dusa McDuff, Yong-Geun Oh, Yuli B. Rudyak, Matthias Schwarz and Claude Viterbo for sending him their preprints on Floer homology and the Arnold conjecture. Finally, he is also grateful to referees for their many very good improvement suggestions.

**References**


Received November 20, 2001 and revised June 6, 2003. The author was supported by the NNSF 19971045 and 10371007 of China.

Department of Mathematics
Beijing Normal University
Beijing 100875
P.R. China
E-mail address: gclu@bnu.edu.cn
TOPOLOGY OF THE MODULI OF REPRESENTATIONS WITH BOREL MOLD

KAZUNORI NAKAMOTO AND TAKESHI TORII

We give descriptions of the moduli of representations with Borel mold for free monoids as fibre bundles over the configuration spaces. By using the associated Serre spectral sequences, we study the cohomology rings of the moduli. Also we calculate the virtual Hodge polynomials of them.

1. Introduction.

A representation for a group or a monoid is called a representation with Borel mold if it can be normalized to a representation in upper triangular matrices whose image of the group or monoid generates the algebra of upper triangular matrices. In [Na2] the moduli of representations with Borel mold has been constructed for each group or monoid. The moduli of representations with Borel mold has simpler structure than the moduli of absolutely irreducible representations constructed in [Na1]. In the present paper, for the free monoid case we describe the moduli of representations with Borel mold explicitly, and calculate its cohomology ring.

The moduli of representations with Borel mold has a fibre bundle structure over the configuration space of the affine space, and hence its cohomology ring can be calculated. We also calculate the virtual Hodge polynomial of the moduli of representations with Borel mold, which will be used for calculating the virtual Poincaré polynomial of the moduli of absolutely irreducible representations of degree 2 for the free monoid case in [Na3]. By calculating the cohomology ring of the moduli, we can consider characteristic classes for representations with Borel mold on a scheme. The construction of characteristic classes and its application will be presented in other papers.

By global representation theory we understand theory of representations on (arbitrary) schemes. The global representation theory is geometric rather than the local representation theory, that is, the representation theory over fields or local rings. For example, each representation of degree $n$ with Borel mold for a group (or a monoid) $\Gamma$ on a scheme $X$ has a unique $\Gamma$-invariant complete flag of $O_X^n$ (see [Na2]). The $\Gamma$-invariant complete flag is not always trivial on $X$, although if $X$ is the spectrum of a field or a local ring, then the flag is trivial. Non-triviality of the $\Gamma$-invariant flag is an interesting feature of the theory of representations over schemes.
As above, the global representation theory has a geometric aspect. For developing the global representation theory, in particular, the theory of representations with Borel mold, we need to consider topology of the moduli of representations with Borel mold. If we intend to construct characteristic classes of representations with Borel mold on schemes (which seems to be an important tool for the global representation theory in the future), then we have to calculate the (ordinary) cohomology of the moduli. That is our main motivation. In this article, we deal with only the free monoid case. The free monoid case is a fundamental case for considering the moduli of representations with Borel mold.

Let us go into details on our main results. There is a 1-1 correspondence between the representations with Borel mold of the free monoid of rank $m$ and the $m$-matrices of size $n \times n$ which generate the algebra of upper triangular matrices. We study the condition for $m$-upper triangular matrices to generate the algebra and it gives us a description of the moduli $\text{Ch}_n^m(B)$ of representations with Borel mold for the free monoid of rank $m$ as a fibre bundle over the configuration space $F_n(A_m \mathbb{Z})$ of the affine space $A_m \mathbb{Z}$.

**Theorem 1.1** (Proposition 3.8). The moduli $\text{Ch}_n^m(B)$ of representations with Borel mold is a fibre bundle over the configuration space $F_n(A_m \mathbb{Z})$ of the affine space $A_m \mathbb{Z}$ with fibre $(\mathbb{P}_Z^{m-2})^{n-1} \times (\mathbb{A}^{m-1}_Z)^{(n-2)(n-1)/2}$ with respect to Zariski topology.

Thereby we can calculate the cohomology rings of $\text{Ch}_n^m(B)$ and related varieties which are regarded as algebraic schemes over $\mathbb{C}$ by tensoring with $\mathbb{C}$. The description of $\text{Ch}_n^m(B)$ as a fibre bundle gives us the Serre spectral sequence converging to the cohomology ring of $\text{Ch}_n^m(B)$. The structure of the cohomology ring of the configuration space $F_n(\mathbb{C}^m)$ is well-known (cf. [Co1] and [Co2]). Then it is easy to show that the spectral sequence collapses from the $E_2$-term.

**Theorem 1.2** (Theorem 5.2). The cohomology ring of $\text{Ch}_n^m(B)$ is given by

$$H^*(\text{Ch}_n^m(B)) \cong H^*(F_n(\mathbb{C}^m)) \otimes \mathbb{Z}[t_1, \ldots, t_{n-1}]/(t_1^{m-1}, \ldots, t_{n-1}^{m-1}),$$

where the degree $|t_j| = 2$ for $1 \leq j \leq n - 1$.

From Deligne’s mixed Hodge theory ([De1] and [De2]), we have an invariant of algebraic varieties over $\mathbb{C}$ called the virtual Hodge polynomial which is a generalization of the Hodge polynomial for smooth projective varieties over $\mathbb{C}$. The virtual Hodge polynomial has a good property for fibre bundles with respect to Zariski topology. We calculate the virtual Hodge polynomials of $\text{Ch}_n^m(B)$ and related varieties.
Theorem 1.3 (Proposition 7.8). The virtual Hodge polynomial of the moduli \( Ch_n(m)_B \) is given by

\[
H(Ch_n(m)_B) = \frac{(z^{m-1} - 1)^{n-1}}{(z - 1)^{n-1}} z^{(m-1)(n-2)(n-1)/2} \prod_{k=0}^{n-1} (z^m - k).
\]

The organization of this paper is as follows: In §2 we review the moduli of representations with Borel mold. In §3 we give descriptions of the moduli schemes \( B_n(m)_B, Ch_n(m)_B \) and \( \text{Rep}_n(m)_B \). We show that \( B_n(m)_B \) and \( Ch_n(m)_B \) are fibred bundles over the configuration space \( F_n(\mathbb{A}^n_Z) \), and \( \text{Rep}_n(m)_B \) is a fibre bundle over the flag scheme \( \text{Flag}(\mathbb{A}^n_Z) \) with respect to Zariski topology. From the descriptions as fibre bundles, we study the associated Serre spectral sequences and calculate the cohomology rings of \( B_n(m)_B, Ch_n(m)_B \) and \( \text{Rep}_n(m)_B \) in §§4, 5, 6. In §7 we calculate the virtual Hodge polynomials of \( B_n(m)_B, Ch_n(m)_B \) and \( \text{Rep}_n(m)_B \). In §8 we define \( B_n(\infty)_B, Ch_n(\infty)_B \) and \( \text{Rep}_n(\infty)_B \) to be the homotopy direct limits of natural inclusions respectively, and study the homotopy types and the cohomology rings of them.


In this section, we make a survey of the moduli of representations with Borel mold. We use [Na2] as our main reference.

2.1. Representations with Borel mold. Let \( \Gamma \) be a group or a monoid. Let \( X \) be a scheme. By a representation of degree \( n \) for \( \Gamma \) on \( X \) we understand a group (resp. monoid) homomorphism \( \Gamma \to \text{GL}_n(\Gamma(X, \mathcal{O}_X)) \) (resp. \( \Gamma \to M_n(\Gamma(X, \mathcal{O}_X)) \)).

For two representations \( \rho, \rho' \) of degree \( n \) for \( \Gamma \) on \( X \), we say that \( \rho \) and \( \rho' \) are equivalent (or \( \rho \sim \rho' \)) if there exists a \( \Gamma(X, \mathcal{O}_X) \)-algebra isomorphism \( \sigma : M_n(\Gamma(X, \mathcal{O}_X)) \to M_n(\Gamma(X, \mathcal{O}_X)) \) such that \( \sigma(\rho(\gamma)) = \rho'(\gamma) \) for each \( \gamma \in \Gamma \).

By a mold of degree \( n \) on a scheme \( X \) we understand a subsheaf of \( \mathcal{O}_X \)-algebras of \( M_n(\mathcal{O}_X) \) which is also a subbundle of \( M_n(\mathcal{O}_X) \). By two molds \( \mathcal{A} \) and \( \mathcal{B} \) of degree \( n \) on \( X \), we say that \( \mathcal{A} \) and \( \mathcal{B} \) are locally equivalent if there exist an open covering \( X = \bigcup_{i \in I} U_i \) and \( P_i \in \text{GL}_n(\Gamma(U_i, \mathcal{O}_X)) \) such that \( P_i^{-1}(\mathcal{A} |_{U_i}) P_i = \mathcal{B} |_{U_i} \). We define the mold \( B_n \) on \( \text{Spec} \mathbb{Z} \) by \( B_n := \{ (b_{ij}) \in M_n(\mathbb{Z}) \mid b_{ij} = 0 \text{ for each } i > j \} \). For a mold \( \mathcal{A} \) of degree \( n \) on \( X \) we say that \( \mathcal{A} \) is a Borel mold of degree \( n \) if \( \mathcal{A} \) and \( B_n \otimes_{\mathbb{Z}} \mathcal{O}_X \) are locally equivalent.

Under the above preparations, we introduce the notion of representations with Borel mold.

Definition 2.1. For a representation \( \rho \) of degree \( n \) for a group (or a monoid) \( \Gamma \) on a scheme \( X \) we say that \( \rho \) is a representation with Borel mold if the subsheaf \( \mathcal{O}_X[\rho(\Gamma)] \) of \( M_n(\mathcal{O}_X) \) generated by \( \rho(\Gamma) \) is a Borel mold.
2.2. Review of the moduli of representations with Borel mold.

Let $\Gamma$ be a group or a monoid. The following functor is representable by an affine scheme:

$$\text{Rep}_n(\Gamma) : (\text{Sch})^o \to (\text{Sets})$$

$$X \mapsto \{\text{representations of degree } n \text{ for } \Gamma \text{ on } X\}.$$ 

The affine scheme $\text{Rep}_n(\Gamma)$ is called the representation variety of degree $n$ for $\Gamma$.

**Definition 2.2.** We define the locally closed subscheme $\text{Rep}_n(\Gamma)_B$ of the affine scheme $\text{Rep}_n(\Gamma)$ which represents the functor

$$\text{Rep}_n(\Gamma)_B : (\text{Sch})^o \to (\text{Sets})$$

$$X \mapsto \left\{ \rho \in \text{Rep}_n(\Gamma)(X) \mid \rho : \text{representation with Borel mold} \right\}.$$ 

**Definition 2.3.** We define the closed subscheme $B_n(\Gamma)$ of $\text{Rep}_n(\Gamma)$ which represents the functor

$$B_n(\Gamma) : (\text{Sch})^o \to (\text{Sets})$$

$$X \mapsto \left\{ \rho \in \text{Rep}_n(\Gamma)(X) \mid \begin{array}{l} \text{the } (i,j)\text{-entry of } \\
\text{the } \rho(\gamma) = 0 \text{ for each } i > j \\
\text{and for each } \gamma \in \Gamma \end{array} \right\}.$$ 

We also define the open subscheme $B_n(\Gamma)_B$ of $B_n(\Gamma)$ by $B_n(\Gamma)_B := B_n(\Gamma) \cap \text{Rep}_n(\Gamma)_B$.

The group scheme $\text{PGL}_n$ acts on the schemes $\text{Rep}_n(\Gamma)$ and $\text{Rep}_n(\Gamma)_B$ by $\rho \mapsto P^{-1}\rho P$. Let $B_n$ be the closed subgroup scheme of $\text{PGL}_n$ defined by $B_n := \{(b_{ij}) \in \text{PGL}_n \mid b_{ij} = 0 \text{ for each } i > j\}$. The group scheme $B_n$ acts on the schemes $B_n(\Gamma)$ and $B_n(\Gamma)_B$ by $\rho \mapsto b\rho b^{-1}$.

We define two group actions on $B_n(\Gamma)_B \times \text{PGL}_n$: One is the action of $\text{PGL}_n$ defined by $(\rho, P) \mapsto (\rho, PQ)$, and the other is one of $B_n$ defined by $(\rho, P) \mapsto (b\rho b^{-1}, bP)$. Defining the morphism $B_n(\Gamma)_B \times \text{PGL}_n \to \text{Rep}_n(\Gamma)_B$ by $(\rho, P) \mapsto P^{-1}\rho P$, we obtain the following diagram which is a fibre product:

$$\begin{array}{ccc}
B_n(\Gamma)_B \times \text{PGL}_n & \to & \text{Rep}_n(\Gamma)_B \\
\downarrow & & \downarrow \\
B_n(\Gamma)_B & \to & \text{Ch}_n(\Gamma)_B.
\end{array}$$

We denote the universal geometric quotient $B_n(\Gamma)_B/B_n = \text{Rep}_n(\Gamma)_B/\text{PGL}_n$ by $\text{Ch}_n(\Gamma)_B$ (the existence of the universal geometric quotient has been proved in $[\text{Na2}]$). The morphism $B_n(\Gamma)_B \times \text{PGL}_n \to B_n(\Gamma)_B$ is the first projection. The two down arrows are $\text{PGL}_n$-principal fibre bundles, and the two right arrows are $B_n$-principal fibre bundles.

Under the above situation, we have the following theorem:
Theorem 2.4 ([Na2]). The scheme $\text{Ch}_n(\Gamma)_B$ represents the sheafification of the following functor with respect to Zariski topology:

$$\mathcal{E}_q B_n(\Gamma) : (\text{Sch})^o \to (\text{Sets})$$

$$X \mapsto \text{Rep}_n(\Gamma)_B(X)/\sim.$$  

In other words, the scheme $\text{Ch}_n(\Gamma)_B$ is the moduli of representations with Borel mold.

By introducing the following notation, we end this section:

Notation 2.5. Let $\Upsilon_m$ be the free monoid of rank $m$. For $\text{Rep}_n(\Upsilon_m)_B$, $B_n(\Upsilon_m)_B$, $\text{Ch}_n(\Upsilon_m)_B$, we also write $\text{Rep}_n(m)_B$, $B_n(m)_B$, $\text{Ch}_n(m)_B$, respectively. These are schemes over $\mathbb{Z}$, however in §4–§8 we use these notations for $\text{Rep}_n(m)_B \otimes \mathbb{Z} \mathbb{C}$, $B_n(m)_B \otimes \mathbb{Z} \mathbb{C}$, and $\text{Ch}_n(m)_B \otimes \mathbb{Z} \mathbb{C}$, respectively.

3. Description of the moduli.

In this section, we describe the moduli of representations with Borel mold of degree $n$ for free monoids by using the configuration spaces. Considering the 1-1 correspondence between representations of the free monoid $\Upsilon_m$ and $m$ matrices of size $n \times n$, we see that $B_n(m)_B$ is isomorphic to

$$\{ (A_1, \ldots, A_m) \mid A_1, \ldots, A_m \text{ generate the algebra of upper triangular matrices} \}.$$  

Hence we will investigate the latter space.

3.1. Preliminaries. In this subsection, we study the condition that $m$ upper triangular matrices generate the algebra of upper triangular matrices.

Let $k$ be a field. We define the $k$-algebra $B_n(k)$ by

$$B_n(k) := \{ (a_{ij}) \in M_n(k) \mid a_{ij} = 0 \text{ for each } i > j \}.$$  

For $n \geq 2$ and $1 \leq i \neq j \leq n$, we define the $k$-linear map

$$p_{ij} : B_n(k) \to B_2(k)$$

$$(a_{st})_{1 \leq s,t \leq n} \mapsto (a_{st})_{s,t=i,j}.$$  

We also define the $k$-algebra homomorphism $\phi_n$ by

$$\phi_n : B_n(k) \to k^n$$

$$(a_{st})_{1 \leq s,t \leq n} \mapsto (a_{11}, a_{22}, \ldots, a_{nn}).$$

Lemma 3.1. For a $k$-subalgebra $\mathcal{A} \subseteq B_2(k)$, it is equal to $B_2(k)$ if and only if $\mathcal{A}$ is a non-commutative algebra.

Proof. Easy. \qed

Lemma 3.2. For a $k$-subalgebra $\mathcal{A} \subseteq B_n(k)$ with $n \geq 3$, $\mathcal{A} = B_n(k)$ if and only if $\phi_n|_\mathcal{A}$ is surjective and $p_{i,i+1}|_\mathcal{A}$ is surjective for each $1 \leq i \leq n - 1$. 
Proof. The “only if” part is obvious. Let us show the “if” part. Put \( J := \text{Ker}(\phi_n|_A) = \{(a_{ij}) \in A \mid a_{11} = a_{22} = \cdots = a_{nn} = 0\} \). First we claim that for each \( 1 \leq i \leq n - 1 \) there exists \( P_i = (a_{st}^i) \in J \) such that \( a_{i,i+1}^i = 1 \) and \( a_{j,j+1}^i = 0 \) for each \( j \neq i \). Since \( p_{i,i+1} | A \) is surjective, there exists \( P' = (b_{st}) \in A \) such that \( b_{ij} = b_{i+1,i+1} = 0 \) and \( b_{i,i+1} = 1 \). By the surjectivity of \( \phi_n|_A \) we also have \( Q := (c_{st}) \in A \) such that \( c_{ii} = 1 \) and \( c_{jj} = 0 \) for each \( j \neq i \). Then the matrix \( P := Q^2P' \) is what we want.

Next let \( P_i \) be as above. For \( 1 \leq i < j \leq n \), put \( X_{ij} := P_i P_{i+1} \cdots P_{j-1} \). Then \( X_{ij} \)'s form a basis of \( \text{Ker}(\phi_n : B_n(k) \to k^n) \). Using the surjectivity of \( \phi_n|_A \) again, we have \( A = B_n(k) \). \( \square \)

**Lemma 3.3.** Let \( k \) be a field. Let \( v_1, v_2, \ldots, v_m \) be \( m \) elements of the \( k \)-algebra \( k^n \). Then \( v_1, v_2, \ldots, v_m \) generate \( k^n \) as a \( k \)-algebra if and only if for each \( 1 \leq i \neq j \leq n \) there exists \( v_{\ell} \) whose \( i \)-th entry and \( j \)-th entry are distinct.

**Proof.** Let \( A \) be the subalgebra of \( k^n \) generated by \( v_1, v_2, \ldots, v_m \). First we show the “if” part. From the assumption, for each \( 1 \leq i \leq n \) and for \( j \neq i \) we have \( w_{ij} \in A \) whose \( i \)-th entry and \( j \)-th entry are 1 and 0, respectively. Since \( \prod_{j \neq i} w_{ij} = e_i = (0, \ldots, 1, \ldots, 0) \), we see that \( A = k^n \).

Next we show that the “only if” part. Suppose that \( A = k^n \) and that there exist \( 1 \leq i \neq j \leq n \) such that the \( i \)-th entry and the \( j \)-th entry of any \( v_{\ell} \) coincide. Then \( A \) is contained in \( \{(a_1, \ldots, a_n) \in k^n \mid a_i = a_j\} \), which is a contradiction. \( \square \)

Let \( A_1, \ldots, A_m \) be \( m \) upper triangular matrices of \( M_n(k) \). Put

\[
A_i = \begin{pmatrix}
a(i)_{11} & a(i)_{12} & a(i)_{13} & \cdots & a(i)_{1n} \\
0 & a(i)_{22} & a(i)_{23} & \cdots & a(i)_{2n} \\
0 & 0 & a(i)_{33} & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & a(i)_{nn}
\end{pmatrix}
\]

We define the vectors \( w_{i,j} \) for \( 1 \leq i \leq j \leq n \) by

\[
w_{i,j} = (a(1)_{i,j}, a(2)_{i,j}, \ldots, a(m)_{i,j}).
\]

Under this situation, we have the following proposition:

**Proposition 3.4.** For \( m \) upper triangular matrices \( A_1, \ldots, A_m \) of \( M_n(k) \), they generate \( B_n(k) \) if and only if \( w_{11}, w_{22}, \ldots, w_{nn} \) are distinct vectors, and two vectors \( w_{ii} - w_{i+1,i+1} \) and \( w_{i,i+1} \) are linearly independent for \( 1 \leq i \leq n - 1 \).

**Proof.** By Lemma 3.2 we see that \( m \) upper triangular matrices \( A_1, \ldots, A_m \) generate \( B_n(k) \) if and only if \( \phi_n(A_1), \ldots, \phi_n(A_m) \) generate \( k^n \) and \( p_{i,i+1}(A_1), \ldots, p_{i,i+1}(A_m) \) are linearly independent for \( 1 \leq i \leq n - 1 \).
\[ \cdots, p_{i,i+1}(A_m) \text{ generate } B_2(k) \text{ for each } 1 \leq i \leq n - 1. \] From Lemma 3.3, \( \phi_n(A_1), \ldots, \phi_n(A_m) \) generate \( k^n \) if and only if \( w_{11}, w_{22}, \ldots, w_{nn} \) are distinct vectors. By using Lemma 3.1 we easily check that \( p_{i,i+1}(A_1), \ldots, p_{i,i+1}(A_m) \) generate \( B_2(k) \) if and only if \( w_{ii} - w_{i+1,i+1} \) and \( w_{i,i+1} \) are linearly independent. Hence we can prove the statement. \( \square \)

3.2. Description of \( B_n(m)_B \). In this subsection, we describe \( B_n(m)_B \) explicitly by using the configuration space of the affine space. Note that \( B_n(m)_B \) is the scheme of \( m \) upper triangular \( n \times n \) matrices which generate the algebra of upper triangular matrices.

**Definition 3.5.** We define the configuration space \( F_n(X) \) of a scheme \( X \) by

\[ F_n(X) := \{ (p_1, p_2, \ldots, p_n) \in X^n \mid p_i \neq p_j \text{ for } i \neq j \}. \]

For example, we denote by \( F_n(A^n_m) \) the configuration space of ordered distinct \( n \)-points in \( A^n_m \).

Let \( A_1, A_2, \ldots, A_m, w_{i,j} \) be as in (1) and (2). We define the morphism \( \Phi_{n,m} : B_n(m)_B \rightarrow F_n(A^n_m) \) by \( (A_1, \ldots, A_m) \mapsto (w_{11}, w_{22}, \ldots, w_{nn}) \). The morphism \( \Phi_{n,m} \) is well-defined by Proposition 3.4. Let us denote \( B_n(Y_m) \) by \( B_n(m) \). We define the isomorphism \( \Xi_{n,m} : B_n(m) \rightarrow (A^n_m)^n \times (A^n_m)^{n-1} \times (A^n_m)^{(n-2)(n-1)/2} \) by

\[ (A_1, \ldots, A_m) \mapsto ((w_{11}, w_{22}, \ldots, w_{nn}), (w_{12}, w_{23}, \ldots, w_{n-1,n}), (w_{i,j})_{|i-j| \geq 2}). \]

Under these preparations, we obtain:

**Proposition 3.6.** Let \( n \) be an integer with \( n \geq 2 \). The morphism \( \Phi_{n,m} : B_n(m)_B \rightarrow F_n(A^n_m) \) is a fibre bundle with fibre \( (A^n_m \setminus A^1_n)^{n-1} \times (A^n_m)^{(n-2)(n-1)/2} \). More precisely, there exists a Zariski open covering \( F_n(A^n_m) = \cup_{i \in I} U_i \) such that \( \Phi_{n,m}^{-1}(U_i) \cong U_i \times (A^n_m \setminus A^1_n)^{n-1} \times (A^n_m)^{(n-2)(n-1)/2} \) and the structure group is \( G := \underbrace{G_0 \times \cdots \times G_0}_{n-1} \), where

\[ G_0 := \left\{ \begin{pmatrix} 1 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{pmatrix} \in \text{GL}_m \right\}. \]

**Proof.** Set \( A^n_m \setminus A^1_n := \{ (t_1, t_2, \ldots, t_m) \in A^n_m \mid (t_2, \ldots, t_m) \neq (0, \ldots, 0) \} \). Let \( G_0 \) act on \( A^n_m \setminus A^1_n \) by \( \tau(t_1, t_2, \ldots, t_m) \mapsto \Lambda \tau(t_1, t_2, \ldots, t_m) \). Then we define the action of \( G \) on \( (A^n_m \setminus A^1_n)^{n-1} \times (A^n_m)^{(n-2)(n-1)/2} \) by

\[ ((z_1, \ldots, z_{n-1}), (w_{ij})_{|i-j| \geq 2}) \mapsto ((A_1 z_1, \ldots, A_{n-1} z_{n-1}), (w_{ij})_{|i-j| \geq 2}) \]

for \( (A_1, \ldots, A_{n-1}) \in G = G_0 \times \cdots \times G_0 \).
For each point of $F_n(\mathbb{A}^m_Z)$, we have a neighbourhood $U$ such that there exist $n-1$ bases

$$\{v_1 - v_2, u(1)_{2}, \ldots, u(1)_m\},$$
$$\{v_2 - v_3, u(2)_{2}, \ldots, u(2)_m\},$$
$$\ldots$$
$$\{v_{n-1} - v_n, u(n-1)_{2}, \ldots, u(n-1)_m\}$$

of $U$-valued points of $\mathbb{A}^m_Z$ for $(v_1, \ldots, v_n) \in U$. We define the isomorphism $U \times (\mathbb{A}^m_Z \setminus \mathbb{A}^1_Z)^{n-1} \times (\mathbb{A}^m_Z)^{(n-2)(n-1)/2} \to \Phi_{n,m}(U)$ by

$$((v_1, \ldots, v_n), (t(i)_1, \ldots, t(i)_m), (w_{ij})_{i-j \geq 2}) \mapsto \Xi_{n,m}^{-1}((v_1, \ldots, v_n), (w_{12}, \ldots, w_{n-1,n})), (w_{ij})_{i-j \geq 2},$$

where $w_{i,i+1} := t(i)_1(v_i-v_{i+1}) + t(i)_2u(i)_2 + \cdots + t(i)_mu(i)_m$ for $1 \leq i \leq n-1$. Thus we easily see that $\Phi_{n,m}$ is a fiber bundle with the structure group $G$. □

**Remark 3.7.** We remark that if $m = 1$, then $B_n(1)_B$ is empty. Hence $\text{Ch}_0(1)_B$ and $\text{Rep}_n(1)_B$ are also empty. If $n = 1$, then $\text{Rep}_1(m) = B_1(m)_B = \text{Ch}_1(m)_B = \mathbb{A}^m_Z$. Therefore in the sequel we assume that $n, m \geq 2$.

**3.3. Description of $\text{Ch}_n(m)_B$.** In this subsection, we describe the moduli of representations with Borel mold $\text{Ch}_n(m)_B$ explicitly.

The morphism $\Phi_{n,m} : B_n(m)_B \to F_n(\mathbb{A}^m_Z)$ is $B_n$-equivariant. Here the group scheme $B_n$ acts on $F_n(\mathbb{A}^m_Z)$ trivially. Hence $\Phi_{n,m}$ induces $\Psi_{n,m} : \text{Ch}_n(m)_B \to F_n(\mathbb{A}^m_Z)$. For each point of $F_n(\mathbb{A}^m_Z)$, we take an open neighbourhood $U$ as in the proof of Proposition 3.6.

Let us consider the action of $B_n$ on $\Phi_{n,m}^{-1}(U) \cong U \times (\mathbb{A}^m_Z \setminus \mathbb{A}^1_Z)^{n-1} \times \mathbb{A}^m_Z^{(n-2)(n-1)/2}$. Let $x = ((v_1, \ldots, v_n), (t(i)_1, \ldots, t(i)_m), (w_{ij})_{i-j \geq 2}) \in U \times (\mathbb{A}^m_Z \setminus \mathbb{A}^1_Z)^{n-1} \times \mathbb{A}^m_Z^{(n-2)(n-1)/2}$. For $B = (b_{ij}) \in B_n$, set $B^{-1} = (b'_{ij})$. We denote $B \cdot x$ by $((v'_1, \ldots, v'_n), (t'(i)_1, \ldots, t'(i)_m), (w'_{ij})_{i-j \geq 2})$. Then we have

$$v'_i = v_i,$$
$$t'(i)_1 = -\frac{b_{i,i+1}}{b_{i+1,i+1}}t(i)_1 + \frac{b_{ii}}{b_{i,i+1}}t(i)_1,$$
$$t'(i)_2 = \frac{b_{ii}}{b_{i+1,i+1}}t(i)_2,$$
$$\ldots$$
$$t'(i)_m = \frac{b_{ii}}{b_{i+1,i+1}}t(i)_m,$$
$$w'_{ij} = \sum_{k \leq t \leq j} b_{ik}w_{kt}b'_{ij}.$$
By calculating $w'_{ij}$, we have

$$w'_{ij} = b_{ii} w_{ij} b'_{jj} + \sum_{i \leq k \leq j} b_{ik} w_{kk} b'_{kj} + \text{(the other terms)}$$

$$= b_{ii} w_{ij} b'_{jj} - b_{ij} b'_{jj} (w_{j-1j-1} - w_{jj})$$
$$- (b_{i,j-1} b'_{j-1j} + b_{ij} b'_{jj})(w_{j-2j-2} - w_{j-1j-1})$$
$$- (b_{i,j-2} b'_{j-2j} + b_{i,j-1} b'_{j-1j} + b_{ij} b'_{jj})(w_{j-3j-3} - w_{j-2j-2}) - \cdots$$
$$- (b_{i,i+1} b'_{i+1j} + \cdots + b_{ij} b'_{jj})(w_{ii} - w_{i+1i+1}) + \left( \sum_{i \leq k \leq j} b_{ik} b'_{kj} \right) w_{ii}$$
$$+(\text{the other terms})$$

$$= b_{ii} w_{ij} b'_{jj} - b_{ij} b'_{jj} (v_{j-1} - v_j) - (b_{i,j-1} b'_{j-1j} + b_{ij} b'_{jj})(v_{j-2} - v_{j-1})$$
$$- (b_{i,j-2} b'_{j-2j} + b_{i,j-1} b'_{j-1j} + b_{ij} b'_{jj})(v_{j-3} - v_{j-2}) - \cdots$$
$$- (b_{i,i+1} b'_{i+1j} + \cdots + b_{ij} b'_{jj})(v_i - v_{i+1}) + \text{(the other terms)}.$$  

Here we used the equality $\sum_{i \leq k \leq j} b_{ik} b'_{kj} = \delta_{ij} = 0$ and we denoted $v_k$ by $w_{kk}$.

We define a morphism $\Phi_{n,m}^{-1}(U) \cong U \times (\mathbb{A}_Z^n \setminus \mathbb{A}_Z^1)^{n-1} \times \mathbb{A}_Z^{m(n-2)(n-1)/2} \rightarrow U \times (\mathbb{P}^{m-2})^{n-1} \times (\mathbb{A}_Z^{m-1})^{(n-2)(n-1)/2}$ by

$$(v_1, \ldots, v_n), (t(i)_1, \ldots, t(i)_m)_{1 \leq i \leq n-1}, (w_{ij})_{|j-j| \geq 2} \mapsto ((v_1, \ldots, v_n), (t(i)_1, \ldots, t(i)_m)_{1 \leq i \leq n-1}, (\overline{w}_{ij})_{|i-j| \geq 2},$$
where $\overline{w}_{ij} \in \mathbb{A}_Z^{m-1} = \langle u_{ij}(2), \ldots, u_{ij}(m) \rangle \subset \mathbb{A}_Z^m = \langle (v_{j-1} - v_j), u_{ij}(2), \ldots, u_{ij}(m) \rangle$ is defined as follows: We take $(v_{j-1} - v_j), u_{ij}(2), \ldots, u_{ij}(m)$ as a basis of $U \times \mathbb{A}_Z^n$ over $U$. The above calculation follows that $w'_{ij} = b_{ii} w_{ij} b'_{jj} - b_{ij} b'_{jj} (v_{j-1} - v_j) + \cdots$, and hence by choosing suitable $b_{ij}$, we can assume that $w'_{ij} \in \langle u_{ij}(2), \ldots, u_{ij}(m) \rangle$. Then we put $\overline{w}_{ij} = w'_{ij}$. Let $B_n$ act on $U \times (\mathbb{P}^{m-2})^{n-1} \times (\mathbb{A}_Z^{m-1})^{(n-2)(n-1)/2}$ trivially. Then the above morphism is $B_n$-equivariant. Since the pull-back $\Psi_{n,m}^{-1}(U) \subseteq \text{Ch}_n(m)_B$ of $U$ is a universal geometric quotient of $\Phi_{n,m}^{-1}(U)$ by $B_n$, the morphism $\Phi_{n,m}^{-1}(U) \rightarrow U \times (\mathbb{P}^{m-2})^{n-1} \times (\mathbb{A}_Z^{m-1})^{(n-2)(n-1)/2}$ induces a morphism $\varphi : \Psi_{n,m}^{-1}(U) \rightarrow U \times (\mathbb{P}^{m-2})^{n-1} \times (\mathbb{A}_Z^{m-1})^{(n-2)(n-1)/2}$.

We can easily check that $\varphi$ gives a bijection between geometric points. The schemes $\Psi_{n,m}(U)$ and $U \times (\mathbb{P}^{m-2})^{n-1} \times (\mathbb{A}_Z^{m-1})^{(n-2)(n-1)/2}$ are smooth over $\mathbb{Z}$. Because $\varphi$ is birational, it is an isomorphism by Zariski’s Main Theorem. Therefore we have:

**Proposition 3.8.** The morphism $\Psi_{n,m} : \text{Ch}_n(m)_B \rightarrow F_n(\mathbb{A}_Z^m)$ is a fibre bundle with fibre $(\mathbb{P}^{m-2})^{n-1} \times (\mathbb{A}_Z^{m-1})^{(n-2)(n-1)/2}$. 
Remark 3.9. The morphism $\Psi_{n,m} : \text{Ch}_n(m)_B \to F_n(\mathbb{A}_Z^m)$ can be interpreted as follows: Let $\text{Ch}_1(m)$ be the moduli of characters for the free monoid $Y_m$ of rank $m$. The scheme $F_n(\mathbb{A}_Z^m)$ is isomorphic to the configuration space $F_n(\text{Ch}_1(m))$ of $\text{Ch}_1(m)$ defined by $F_n(\text{Ch}_1(m)) := \{(\chi_1, \ldots, \chi_n) \in \text{Ch}_1(m) \mid \chi_i \neq \chi_j \text{ for } i \neq j\}$, since $\text{Ch}_1(m) \cong \mathbb{A}_Z^n$. For $\rho \in B_n(m)_B$ we can define $(\rho_1, \rho_2, \ldots, \rho_{mn}) \in F_n(\text{Ch}_1(m))$. This correspondence induces $\Psi_{n,m} : \text{Ch}_n(m)_B \to F_n(\mathbb{A}_Z^m) \cong F_n(\text{Ch}_1(m))$. The fibre $\Psi_{n,m}^{-1}(\chi_1, \ldots, \chi_n)$ corresponds to the equivalence classes of representations with Borel mold which are extensions of characters $(\chi_1, \ldots, \chi_n)$.

In the case $n = 2$, Proposition 3.8 says that $\Psi_{2,m} : \text{Ch}_2(m)_B \to F_2(\mathbb{A}_Z^m)$ is a fibre bundle with fibre $\mathbb{P}^{m-2}_Z$. In particular, $\text{Ch}_2(2)_B$ is isomorphic to $F_2(\mathbb{A}_Z^2) \cong \mathbb{A}_Z^2 \times (\mathbb{A}_Z^2 \setminus \{0\})$. Let us describe the $\mathbb{P}^{m-2}_Z$-bundle over $F_2(\mathbb{A}_Z^m)$ more precisely.

The configuration space $F_2(\mathbb{A}_Z^m)$ is isomorphic to $\mathbb{A}_Z^m \times (\mathbb{A}_Z^m \setminus \{0\})$ by $(v_1, v_2) \mapsto (v_1, v_1 - v_2)$. We denote by $f$ the composition of morphisms

$$F_2(\mathbb{A}_Z^m) \cong \mathbb{A}_Z^m \times (\mathbb{A}_Z^m \setminus \{0\}) \to \mathbb{A}_Z^m \setminus \{0\} \to \mathbb{P}^{m-1}_Z.$$

Let us consider the short exact sequence

$$0 \to \mathcal{O}_{\mathbb{P}^{m-1}_Z}(-1) \to \mathcal{O}_{\mathbb{P}^{m-1}_Z}^{\oplus m} \to T_{\mathbb{P}^{m-1}_Z}(-1) \to 0$$

on $\mathbb{P}^{m-1}_Z$. Put $\mathcal{E} := f^*T_{\mathbb{P}^{m-1}_Z}(-1)$. The morphism $\Psi_{2,m} : \text{Ch}_2(m)_B \to F_2(\mathbb{A}_Z^m)$ is described as follows:

**Proposition 3.10.** The moduli $\text{Ch}_2(m)_B$ is isomorphic to $\mathcal{P}_{\text{Proj} \mathcal{E}}$ over $F_2(\mathbb{A}_Z^m)$.

**Proof.** Recall that the morphism $\Psi_{2,m} : \text{Ch}_2(m)_B \to F_2(\mathbb{A}_Z^m)$ is given by $[\{A_1, \ldots, A_m\}] \mapsto (w_{11}, w_{22})$, where $A_i, w_{ij}$ are as in (1) and (2). Let us consider the pull-back of (3) by $f \circ \Psi_{2,m}$:

$$0 \to (f \circ \Psi_{2,m})^*\mathcal{O}_{\mathbb{P}^{m-1}_Z}(-1) \xrightarrow{w_{11} - w_{22}} \mathcal{O}_{\text{Ch}_2(m)_B}^{\oplus m} \to \Psi_{2,m}^*\mathcal{E} \to 0.$$

The vectors $w_{12}$ and $w_{11} - w_{22}$ are linearly independent. For $B = (b_{ij}) \in B_2$, the $w_{12}$ vector of $B \cdot (A_1, \ldots, A_m)$ is given by $-b_{12}/b_{22} \cdot (w_{11} - w_{22}) + b_{11}/b_{22} \cdot w_{12}$. From these facts, the vector $w_{12}$ determines a sub-line bundle $\mathcal{L}$ of $\Psi_{2,m}^*\mathcal{E}$. Hence we have the surjection $\Psi_{2,m}^*\mathcal{E} \to \mathcal{L} \to 0$. The surjective homomorphism of algebras $S(\Psi_{2,m}^*\mathcal{E}) \to S(\mathcal{L})$ induces $\text{Ch}_2(m)_B \to \text{Ch}_2(m)_B \times \mathcal{P}_{\text{Proj} \mathcal{E}} \xrightarrow{\text{proj}} \mathcal{P}_{\text{Proj} \mathcal{E}}$. We can easily check that this is an isomorphism.

**3.4. Description of $\text{Rep}_n(m)_B$.** In this subsection, we describe $\text{Rep}_n(m)_B$. 

In §2 we obtained a diagram which is a fibre product:
\[
\begin{array}{ccc}
B_n(m)_B \times \text{PGL}_n & \xrightarrow{f} & \text{Rep}_n(m)_B \\
\downarrow p_1 & & \downarrow \pi \\
B_n(m)_B & \xrightarrow{\pi'} & \text{Ch}_n(m)_B,
\end{array}
\]
where \( f : B_n(m)_B \times \text{PGL}_n \to \text{Rep}_n(m)_B \) is given by \((\rho, P) \mapsto P^{-1} \rho P\) and \( p_1 \) is the first projection. The group scheme \( B_n \) acts on \( B_n(m)_B \times \text{PGL}_n \) by \((\rho, P) \mapsto (Q \rho Q^{-1}, Q P)\). The morphism \( f \) is a \( B_n \)-principal fibre bundle. Hence we conclude that \( B_n(m)_B \times_{B_n} \text{PGL}_n \cong \text{Rep}_n(m)_B \).

The universal representation with Borel mold on \( \text{Rep}_n(m)_B \) induces the action of the free monoid \( \Upsilon_m \) on the trivial bundle \( \mathcal{O}_n^{\oplus n} \). In [Na2] we obtained a unique complete flag \( 0 \subset \mathcal{E}_1 \subset \cdots \subset \mathcal{E}_{n-1} \subset \mathcal{O}_n^{\oplus n} \) such that \( \mathcal{E}_i \) is a unique \( \Upsilon_m \)-invariant subbundle of rank \( i \). Then we get a morphism \( \text{Rep}_n(m)_B \to \text{Flag}(\mathcal{A}_n^m) \) associated to the complete flag, where \( \text{Flag}(\mathcal{A}_n^m) \) is the flag scheme consisting of complete flags of the rank \( n \) trivial bundle.

**Proposition 3.11.** The morphism \( \text{Rep}_n(m)_B \to \text{Flag}(\mathcal{A}_n^m) \) is a fibre bundle with fibre \( B_n(m)_B \).

**Proof.** For each \( x \in \text{Flag}(\mathcal{A}_n^m) \), we can choose an open neighbourhood \( U \) of \( x \) and \( n \) sections \( s_i \) (\( 1 \leq i \leq n \)) of \( \mathcal{O}_U^{\oplus n} \) such that \( \bigoplus_{i=1}^k \mathcal{O}_U \cdot s_i \) is the rank \( k \) subbundle of the universal flag on \( U \). We denote by \( \widetilde{U} \) the inverse image of \( U \) by \( \text{Rep}_n(m)_B \to \text{Flag}(\mathcal{A}_n^m) \). Let \( \mathcal{E}_* \) be the pull-back of the universal flag on \( \widetilde{U} \). Let \( \widetilde{s}_i \) be the pull-back of \( s_i \). Then we define a morphism \( U \times B_n(m)_B \to \widetilde{U} \) by corresponding \( (\mathcal{E}_*, \rho) \) to the representation \( \rho \) with respect to the basis \( \{ \widetilde{s}_i \} \) (not the canonical basis!). We can easily check that \( U \times B_n(m)_B \to \widetilde{U} \) is an isomorphism, which completes the proof. The statement can be also verified by the fact that \( B_n(m)_B \times_{B_n} \text{PGL}_n \cong \text{Rep}_n(m)_B \). \( \square \)

**4. Cohomology of \( B_n(m)_B \).**

In §3.2 we described the scheme \( B_n(m)_B \) over \( \mathbb{Z} \) as a fibre bundle over the configuration space \( F_n(\mathbb{A}_Z^m) \). In the rest of this paper we abbreviate the \( \mathbb{C} \)-valued point of \( B_n(m)_B \) with classical topology to \( B_n(m)_B \). In this section we calculate the cohomology ring of \( B_n(m)_B \) for \( m \geq 2 \) by using the Serre spectral sequence associated with the fibre bundle. For a topological space \( X \), we denote by \( H^q(X) \) the integral cohomology group \( H^q(X; \mathbb{Z}) \).

First, we recall the cohomology ring of the configuration space \( F_n(\mathbb{R}^m) \) (cf. [Co1] and [Co2]). Let \( F_n(\mathbb{R}^m) \) be the configuration space of ordered distinct \( n \)-points in \( \mathbb{R}^m \):
\[
F_n(\mathbb{R}^m) = \{(x_1, \ldots, x_n) \in (\mathbb{R}^m)^n | x_i \neq x_j (i \neq j)\}.
\]
Since \( F_2(\mathbb{R}^m) \) is homotopy equivalent to the \((m-1)\)-sphere \( S^{m-1} \), we have \( H^s(F_2(\mathbb{R}^m)) \cong \Lambda(s) \) where the degree \( |s| = m - 1 \). For \( i \neq j \), we define
the map \( \pi_{ij} : F_n(\mathbb{R}^m) \to F_2(\mathbb{R}^m) \) given by \( \pi_{ij}(x_1, \ldots, x_n) = (x_i, x_j) \). Let 
\( s(i, j) = \pi_{ij}^*(s) \). Then we have \( s(i, j)^2 = 0 \) and \( s(j, i) = (-1)^m s(i, j) \).

**Theorem 4.1** (cf. [Co1] and [Co2]). The cohomology ring of the configuration space \( F_n(\mathbb{R}^m) \) is a graded commutative associative ring generated by 
\( s(i, j) \) for \( 1 \leq i < j \leq n \) with a complete set of relations:
\[
\begin{align*}
\quad s(i, j)^2 &= 0, \\
\quad s(i, k)s(j, k) &= s(i, j)s(j, k) - s(i, j)s(i, k) \quad \text{for} \quad i < j < k.
\end{align*}
\]

By Proposition 3.6, there is a fibre bundle
\[
(4) \quad Y_B \xrightarrow{i} B_n(m)_B \xrightarrow{\Phi_{n,m}} F_n(\mathbb{C}^m),
\]
where the fibre \( Y_B \) is \((\mathbb{C}^m - \mathbb{C}^1)^{n-1} \times \mathbb{C}^{m(n-1)(n-2)/2}\). Since \( Y_B \) is homotopy equivalent to the product of spheres:
\[
Y_B \cong S^{2m-3} \times \cdots \times S^{2m-3},
\]
the cohomology of the fibre \( Y_B \) is given by
\[
H^*(Y_B) \cong \Lambda(s'_1, \ldots, s'_{n-1}),
\]
where the degree of \( s'_j \) is \( 2m - 3 \) for \( j = 1, \ldots, n - 1 \).

**Lemma 4.2.** \( B_n(m)_B \) is \((2m - 4)\)-connected.

**Proof.** Note that the configuration space \( F_n(\mathbb{C}^m) \) is \((2m - 2)\)-connected. Then the lemma follows from the long exact sequence of homotopy groups associated with the fibre bundle (4). \( \square \)

There is a Serre spectral sequence associated with the fibre bundle (4)
\[
E_2^{p,q} = H^p(F_n(\mathbb{C}^m); H^q(Y_B)) \Rightarrow H^{p+q}(B_n(m)_B).
\]
Note that the coefficient system is trivial, since \( F_n(\mathbb{C}^m) \) is \((2m - 2)\)-connected \((m \geq 2)\). Since \( H^*(F_n(\mathbb{C}^m)) \) and \( H^*(Y_B) \) are free over \( \mathbb{Z} \), we have an isomorphism
\[
E_2^{p,q} \cong H^p(F_n(\mathbb{C}^m)) \otimes H^q(Y_B).
\]
By Theorem 4.1, \( H^p(F_n(\mathbb{C}^m)) = 0 \) for \( 1 \leq p \leq 2m - 2 \). Hence this spectral sequence collapses from \( E_2 \)-term. In particular, \( H^*(B_n(m)_B) \) is free over \( \mathbb{Z} \). Since \( i^* : H^{2m-3}(B_n(m)_B) \to H^{2m-3}(Y_B) \) is an isomorphism, there is \( s_j \in H^{2m-3}(B_n(m)_B) \) such that \( i^*(s_j) = s'_j \) for \( j = 1, \ldots, n - 1 \). By using the ring homomorphism \( \Phi_{n,m} : H^*(F_n(\mathbb{C}^m)) \to H^*(B_n(m)_B) \), we regard \( H^*(B_n(m)_B) \) as an algebra over \( H^*(F_n(\mathbb{C}^m)) \).

**Theorem 4.3.** The cohomology ring of \( B_n(m)_B \) is an exterior algebra generated by \( s_1, \ldots, s_{n-1} \) over \( H^*(F_n(\mathbb{C}^m)) \):
\[
H^*(B_n(m)_B) \cong H^*(F_n(\mathbb{C}^m)) \otimes \Lambda(s_1, \ldots, s_{n-1}).
\]
Proof. Since $H^*(B_n(m)_B)$ is free over $\mathbb{Z}$, $s_j^2 = 0$ for $j = 1, \ldots, n - 1$. There is a ring homomorphism $\phi : \Lambda(s_1, \ldots, s_{n-1}) \to H^*(B_n(m)_B)$. Then $\phi$ is injective since $i^* \circ \phi$ is an isomorphism. We consider the following ring homomorphism:

$$
\Phi^*_{n,m} \otimes \phi : H^*(F_n(C^m)) \otimes \Lambda(s_1, \ldots, s_j) \to H^*(B_n(m)_B).
$$

Then it is easy to see that $\Phi^*_{n,m} \otimes \phi$ is an isomorphism. \qed

For $(A_1, \ldots, A_m) \in B_n(m)_B$, we recall that $a(i)_{k,l}$ is the $(k,l)$-entry of the $i$th matrix $A_i$. Then they define a vector $w_{k,l}$ in $C^m$ by $w_{k,l} = (a(1)_{k,l}, \ldots, a(m)_{k,l})$. We set $w_k = w_{k,k} - w_{k+1,k+1}$ for $k = 1, \ldots, n - 1$. Let $B_n(m)'_B$ be the subspace of $B_n(m)_B$ defined as follows:

$$
B_n(m)'_B = \left\{ (A_1, \ldots, A_m) \in B_n(m)_B \midegin{array}{ll}
av(i)_{k,l} = 0 & (1 \leq i \leq m, l > k + 1), \\
(w_k, w_{k+1}) = 0 & (1 \leq k \leq n - 1), \\
|w_{k+1}| = 1 & (1 \leq k \leq n - 1)
\end{array}\right\},
$$

where $(-, -)$ is the standard Hermitian inner product and $| - |$ is the associated norm. Let $T^n$ be the $n$-dimensional torus $S^1 \times \cdots \times S^1$. Then there is a homomorphism from $T^n$ into the diagonal matrices of $B_n(C)$. We denote by $T^n_\mathbb{R}$ the image of this homomorphism. Then $B_n(m)'_B$ is a $T^n_\mathbb{R}$-equivariant subspace of $B_n(m)_B$ where the action of $T^n_\mathbb{R}$ on $B_n(m)_B$ is a restriction of the action of $B_n(C)$. We note that $T^n_\mathbb{R}$ acts on $B_n(m)'_B$ freely. Then the following lemma is easy:

**Lemma 4.4.** $B_n(m)'_B \hookrightarrow B_n(m)_B$ is a $T^n_\mathbb{R}$-equivariant homotopy equivalence.

The map from $B_n(m)'_B$ to $F_n(C^m)$ gives a fibre bundle

$$
Y_B' \longrightarrow B_n(m)'_B \longrightarrow F_n(C^m),
$$

where the fibre $Y_B'$ is the product of spheres:

$$
Y_B' = S^{2m-3} \times \cdots \times S^{2m-3}.
$$

There is a map of fibre bundles from $Y_B' \rightarrow B_n(m)'_B \rightarrow F_n(C^m)$ to $Y_B \rightarrow B_n(m)_B \rightarrow F_n(C^m)$ which induces homotopy equivalences:

$$
Y_B' \underset{\simeq}{\longrightarrow} B_n(m)'_B \underset{\simeq}{\longrightarrow} F_n(C^m)
$$

$$
Y_B \underset{\simeq}{\longrightarrow} B_n(m)_B \underset{\simeq}{\longrightarrow} F_n(C^m).
$$
5. Cohomology of $\text{Ch}_n(m)_B$.

In the rest of this paper we abbreviate the $\mathbb{C}$-valued point of $\text{Ch}_n(m)_B$ with classical topology to $\text{Ch}_n(m)_B$. In §3.3 we obtained a description of the scheme $\text{Ch}_n(m)_B$ over $\mathbb{Z}$ as a fibre bundle over the configuration space $F_n(\mathbb{A}^m_\mathbb{C})$. By using the Serre spectral sequence of the fibre bundle, we calculate the cohomology ring of $\text{Ch}_n(m)_B$ for $m \geq 2$.

The space $\text{Ch}_n(m)_B$ is defined to be the quotient space of $B_n(m)_B$ by the free action of $B_n(\mathbb{C})$. The torus $T_\mathbb{R} \subset B_n(\mathbb{C})$ also acts on $B_n(m)_B$. There is a fibre bundle

$$B_n(\mathbb{C})/T_\mathbb{R} \rightarrow B_n(m)_B/T_\mathbb{R} \rightarrow \text{Ch}_n(m)_B.$$  

Since the fibre $B_n(\mathbb{C})/T_\mathbb{R}$ is contractible, the projection $B_n(m)_B/T_\mathbb{R} \rightarrow \text{Ch}_n(m)_B$ is a weak homotopy equivalence. By Lemma 4.4, there is a $T_\mathbb{R}$-subspace $B_n(m)'_B$ of $B_n(m)_B$ such that the inclusion is a $T_\mathbb{R}$-equivariant homotopy equivalence. Let $\text{Ch}_n(m)'_B$ be the quotient space $B_n(m)'_B/T_\mathbb{R}$.

Hence we have the following lemma:

**Lemma 5.1.** $\text{Ch}_n(m)_B$ is weakly homotopy equivalent to $\text{Ch}_n(m)'_B$.

By Lemma 5.1, the natural map $\text{Ch}_n(m)'_B \rightarrow \text{Ch}_n(m)_B$ induces an isomorphism of cohomology rings. Hence we calculate the cohomology of $\text{Ch}_n(m)'_B$. There is a map from $\text{Ch}_n(m)'_B$ to $F_n(\mathbb{C}^m)$ which gives a fibre bundle

$$Y'_C \xrightarrow{i'} \text{Ch}_n(m)'_B \xrightarrow{\Psi'_{n,m}} F_n(\mathbb{C}^m).$$  

Note that we have a commutative diagram of fibre bundles

$$
\begin{array}{ccc}
Y'_C & \xrightarrow{i'} & \text{Ch}_n(m)'_B \\
\downarrow \cong & & \downarrow \cong \\
Y_C & \xrightarrow{i} & \text{Ch}_n(m)_B \\
\end{array}
\quad \xrightarrow{\Psi_{n,m}} \quad F_n(\mathbb{C}^m)
$$

such that the vertical arrows are weak homotopy equivalences. The fibre $Y'_C$ is the product of complex projective spaces:

$$Y'_C = \mathbb{CP}^{m-2} \times \cdots \times \mathbb{CP}^{m-2}.$$  

Hence we have

$$H^*(Y'_C) \cong \mathbb{Z}[t'_1, \ldots, t'_{n-1}]/(t'_1^{m-1}, \ldots, t'_{n-1}^{m-1}),$$

where the degree of $t'_j$ is 2 for $j = 1, \ldots, n - 1$. There is a Serre spectral sequence associated with the fibre bundle (5)

$$E_2^{p,q} = H^p(F_n(\mathbb{C}^m); H^q(Y'_C)) \Rightarrow H^{p+q}(\text{Ch}_n(m)'_B).$$
The coefficient system is trivial by the same reason as in the case of $B_n(m)_B$. Note that there is an isomorphism

$$E_{2,q}^{p,q} \cong H^p(F_n(C^m)) \otimes H^q(Y'_C),$$

since $H^*(F_n(C^m))$ and $H^*(Y'_C)$ are free over $\mathbb{Z}$. By Theorem 4.1, we have $H^p(F_n(C^m)) = 0$ for $1 \leq p \leq 2m - 2$. Then the homomorphism $i^* : H^q(\text{Ch}_n(m)_B) \to H^q(Y'_C)$ is an isomorphism for $q \leq 2m - 2$. Let $t_j$ be an element of $H^2(\text{Ch}_n(m)_B)$ such that $i^*(t_j) = t'_j$ for $j = 1, \ldots, n - 1$. Then we have $t_j^{m-1} = 0$ for $j = 1, \ldots, n - 1$.

We regard $H^*(\text{Ch}_n(m)_B)$ as an algebra over $H^*(F_n(C^m))$ by using the ring homomorphism $\Psi_{n,m}^* : H^*(F_n(C^m)) \to H^*(\text{Ch}_n(m)_B)$.

**Theorem 5.2.** The cohomology ring of $\text{Ch}_n(m)_B$ is a truncated polynomial algebra generated by $t_j$, $(j = 1, \ldots, n - 1)$ over $H^*(F_n(C^m))$:

$$H^*(\text{Ch}_n(m)_B) \cong H^*(F_n(C^m)) \otimes \mathbb{Z}[t_1, \ldots, t_{n-1}]/(t_1^{m-1}, \ldots, t_{n-1}^{m-1}).$$

**Proof.** By the above argument, we have a ring homomorphism

$$\psi : \mathbb{Z}[t_1, \ldots, t_{n-1}]/(t_1^{m-1}, \ldots, t_{n-1}^{m-1}) \to H^*(\text{Ch}_n(m)_B).$$

Then the ring homomorphism

$$H^*(F_n(C^m)) \otimes \mathbb{Z}[t_1, \ldots, t_{n-1}]/(t_1^{m-1}, \ldots, t_{n-1}^{m-1}) \xrightarrow{\psi_{n,m}^* \otimes \psi} H^*(\text{Ch}_n(m)_B)$$

gives an isomorphism. \qed

**6. Cohomology of $\text{Rep}_n(m)_B$.**

In §3.4 we described the scheme $\text{Rep}_n(m)_B$ over $\mathbb{Z}$ as a fibre bundle over the flag scheme $\text{Flag}(\mathbb{A}_g^2)$. In the following we abbreviate the $\mathbb{C}$-valued points of $\text{Rep}_n(m)_B$ with classical topology to $\text{Rep}_n(m)_B$. In this section we consider the cohomology of $\text{Rep}_n(m)_B$ for $m \geq 2$ by using the Serre spectral sequence associated with the fibre bundle.

First, we recall the cohomology ring of the flag manifold $U(n)/T^n$. We say that a sequence $(L_1, \ldots, L_{n-1})$ of subvector spaces in $\mathbb{C}^n$ is a complete flag if $L_i \subset L_{i+1}$ for $i = 1, \ldots, n - 2$ and $\dim_{\mathbb{C}} L_i = i$ for $i = 1, \ldots, n - 1$. Let $\text{Flag}(\mathbb{C}^n)$ be the set of all complete flags in the vector space $\mathbb{C}^n$. Then $\text{PGL}_n(\mathbb{C})$ acts on $\text{Flag}(\mathbb{C}^n)$ transitively. Let $\mathbb{C}^i$ be the subspace of $\mathbb{C}^n$ spanned by the first $i$ canonical basis vectors for $i = 1, \ldots, n - 1$. Then we see that the stabilizer of the complete flag $(\mathbb{C}^1, \ldots, \mathbb{C}^{n-1})$ is $B_n(\mathbb{C})$. We regard $\text{Flag}(\mathbb{C}^n)$ as a manifold by means of the isomorphism $\text{Flag}(\mathbb{C}^n) \cong \text{PGL}_n(\mathbb{C})/B_n(\mathbb{C})$. Let $U(n)$ be the unitary group of size $n$ and let $T^n$ be a maximal torus of $U(n)$ consisting of the diagonal matrices. Then $U(n)$ also acts on $\text{Flag}(\mathbb{C}^n)$ transitively and the stabilizer group of $(\mathbb{C}^1, \ldots, \mathbb{C}^{n-1})$ is $T^n$. Hence we get an isomorphism $\text{Flag}(\mathbb{C}^n) \cong U(n)/T^n$. Let $\pi_t : T^n \to T^1$
be the $i$th projection for $i = 1, \ldots, n$. Then we have a line bundle $E_i$ over $\text{Flag} (\mathbb{C}^n)$:

$$U(n) \times_{\pi_i} \mathbb{C} \to \text{Flag} (\mathbb{C}^n).$$

We denote by $t_i$ the first Chern class of the line bundle $E_i$:

$$t_i = c_1(E_i) \in H^2(\text{Flag}(\mathbb{C}^n)).$$

Then we have the following well-known lemma:

**Lemma 6.1.** The cohomology ring of $\text{Flag} (\mathbb{C}^n)$ is given by

$$H^*(\text{Flag} (\mathbb{C}^n)) = \mathbb{Z}[t_1, \ldots, t_n]/(c_1, \ldots, c_n),$$

where $c_i$ is the $i$th symmetric function for $i = 1, \ldots, n$.

We note that $H^i(\text{Flag} (\mathbb{C}^n)) = 0$ for $i > n^2 - n$ since $\text{Flag} (\mathbb{C}^n)$ is a closed manifold of real dimension $n^2 - n$.

The space $\text{Rep}_n (m)_B$ is defined as $B_n (m)_B \times \text{PGL}_n (\mathbb{C})$. We note that there is an isomorphism $\text{Rep}_n (m)_B \cong B_n (m)_B \times_{T_{\mathbb{R}}} PU(n)$ where $PU(n)$ is the projective unitary group and $T_{\mathbb{R}}$ is its maximal torus. We define $\text{Rep}_n (m)'_B$ as $B_n (m)'_B \times_{T_{\mathbb{R}}} PU(n)$.

**Lemma 6.2.** There is a homotopy equivalence $\text{Rep}_n (m)_B \simeq \text{Rep}_n (m)'_B$.

**Proof.** This follows from Lemma 4.4. \hfill \Box

By Lemma 6.2, the natural map $\text{Rep}_n (m)'_B \to \text{Rep}_n (m)_B$ induces an isomorphism of cohomology rings. Hence we calculate the cohomology of $\text{Rep}_n (m)_B$. There is a fibre bundle

$$B_n (m)_B \to \text{Rep}_n (m)_B \to \text{Flag} (\mathbb{C}^n).$$

Then we obtain the associated Serre spectral sequence

$$E_2^{p,q} = H^p (\text{Flag}(\mathbb{C}^n)); H^q (\text{Rep}_n (m)_B) \implies H^{p+q} (\text{Rep}_n (m)_B).$$

Since $\text{Flag} (\mathbb{C}^n)$ is simply connected, the coefficient system is trivial. By Theorem 4.3 and Lemma 6.1, the cohomology group of $B_n (m)_B$ and $\text{Flag} (\mathbb{C}^n)$ are free over $\mathbb{Z}$. Hence we have an isomorphism

$$E_2^{p,q} \cong H^p (\text{Flag}(\mathbb{C}^n)) \otimes H^q (\text{Rep}_n (m)_B).$$

We recall that there is a map $B_n (m)'_B \to F_n (\mathbb{C}^m)$ which is a fibre bundle with fibre $Y'_B$.

**Lemma 6.3.** Let $c \in H^*(B_n (m)'_B)$. If $c$ is in the image of the homomorphism $H^*(F_n (\mathbb{C}^m)) \to H^*(B_n (m)'_B)$, then $c$ is a permanent cycle.

**Proof.** This follows from the fact that there is a map $\text{Rep}_n (m)'_B \to F_n (\mathbb{C}^m)$ which factors through $B_n (m)'_B \to F_n (\mathbb{C}^m)$. \hfill \Box

**Corollary 6.4.** The $E_2^{*,*}$ is a spectral sequence of $H^*(F_n (\mathbb{C}^m))$-modules.
Proposition 6.5. If $m > (n^2 - n)/2 + 1$, then the spectral sequence collapses from $E_2$-term. In this case we have

$$H^\ast(\text{Rep}_n(m)_B) \cong H^\ast(F_n(C^m)) \otimes H^\ast(\text{Flag}(C^n)) \otimes \Lambda(s_1, \ldots, s_{n-1})$$

as algebras where the degree of $s_i$ is $2m - 3$ for $i = 1, \ldots, n - 1$.

Proof. Since $B_n(m)_B$ is $(2m - 4)$-connected, we have $d_2 = \cdots = d_{2m - 3} = 0$. Then the proposition follows from the fact that $H^i(\text{Flag}(C^n)) = 0$ for $i > n^2 - n$.

The first nontrivial differential $d_{2m-2}$ is given by

$$d_{2m-2}(s_i) = (t_i - t_{i+1})^{m-1} \text{ for } 1 \leq i < n.$$ 

Let $C$ be a differential graded algebra given by

$$C = \mathbb{Z}[t_1, \ldots, t_n]/(c_1, \ldots, c_n) \otimes \Lambda(s_1, \ldots, s_{n-1}),$$

where the cohomological degree of $t_i$ is 0 for $i = 1, \ldots, n$ and the cohomological degree of $s_i$ is 1 for $i = 1, \ldots, n - 1$. The differential is defined by

$$d(s_i) = (t_i - t_{i+1})^{m-1}, \quad i = 1, \ldots, n - 1.$$ 

We denote by $H(C)$ the cohomology algebra of $C$.

Lemma 6.6. The $E_{2m-1}$-term of the Serre spectral sequence of the fibre bundle $B_n(m)_B \to \text{Rep}_n(m)_B \to F_n(C^m)$ is $H(C) \otimes H^\ast(F_n(C^m))$.

In the rest of this section we calculate the cohomology of $\text{Rep}_n(m)$ for small $n$.

6.1. The Case $n = 2$. If $n = 2$, the flag manifold $\text{Flag}(C^2)$ is the 2-sphere $S^2$ and $PU(2)$ is the real projective space $\mathbb{RP}^3$. We recall that $\text{Rep}_2(2)_B = B_2(2)_B \times T_{\mathbb{R}} PU(2)$. It is easy to see that the action of $T_{\mathbb{R}}$ is free and the quotient map $B_2(2)_B \to B_2(2)_B/T_{\mathbb{R}}$ is identified with the fibre bundle $B_2(2)_B \to F_2(C^2)$. Hence $B_2(2)_B \to F_2(C^2)$ is a principal $T_{\mathbb{R}}$-bundle. Since $T_{\mathbb{R}} \cong S^1$ and $F_2(C^2)$ is 2-connected, the principal bundle is trivial and $B_2(2)_B \cong F_2(C^2) \times T_{\mathbb{R}}$. This implies that $\text{Rep}_2(2)_B \cong F_2(C^2) \times PU(2)$. It is also easy to construct an isomorphism explicitly.

Proposition 6.7. If $n = 2$ and $m = 2$, we have a homotopy equivalence $\text{Rep}_2(2)_B \simeq F_2(C^2) \times \mathbb{RP}^3$. Hence its cohomology ring is given by

$$H^\ast(\text{Rep}_2(2)_B) \cong H^\ast(F_2(C^2)) \otimes H^\ast(\mathbb{RP}^3).$$

If $n = 2$ and $m \geq 3$, we have

$$H^\ast(\text{Rep}_2(m)_B) \cong H^\ast(F_2(C^m)) \otimes H^\ast(\text{Flag}(C^2)) \otimes \Lambda(s)$$

where $|s| = 2m - 3$. 

Proof. The case $m = 2$ follows from Lemma 6.2. The case $m = 3$ follows from Proposition 6.5.

Remark 6.8. There exists a unique $\Upsilon_m$-invariant sub-line bundle $\mathcal{L}_m$ of $\mathcal{O}_{\text{Rep}_2(m)_B}^{\otimes 2}$ on $\text{Rep}_2(m)_B$. The line bundle $\mathcal{L}_m$ is obtained by the pull back of $\mathcal{E}_1$ of the universal flag by $\text{Rep}_2(m)_B \to \text{Flag}(\mathbb{C}^2)$ in §3.4. In the case $m = 2$, we see that $0 \neq c_1(\mathcal{L}_2) \in H^2(\text{Rep}_2(2)_B) \cong \mathbb{Z}/2\mathbb{Z}$ by [Na2] Proposition 4.5. We also see that $\mathcal{L}_2^\otimes 2 \cong \mathcal{O}_{\text{Rep}_2(2)_B}$ by [Na2] Proposition 4.7. In the case $m \geq 3$, we have $c_1(\mathcal{L}_m) = t_1 \in H^2(\text{Rep}_2(m)_B) = H^2(\text{Flag}(\mathbb{C}^2)) \cong \mathbb{Z}$.

6.2. The case $\mathbf{n} = 3$. By Lemma 6.6, the $E_{2m-1}$-term of the Serre spectral sequence of the fibre bundle $B_3(m)'_B \to \text{Rep}_3(m)'_B \to \text{Flag}(\mathbb{C}^3)$ is given by

$$E_{2m-1} \cong H(C) \otimes H^*(\text{Flag}(\mathbb{C}^3)).$$

Then the next nontrivial differential is $d_{4m-5}$. Since $H^*(\text{Flag}(\mathbb{C}^3))$ is concentrated in even degrees, we see that $d_{4m-5}(H(C)) = 0$. Hence we obtain the following proposition:

Proposition 6.9. If $n = 3$, then we have $H^*(\text{Rep}_3(m)_B; k) \cong H^*(C \otimes k) \otimes H^*(\text{Flag}(\mathbb{C}^m); k)$ as $H^*(\text{Flag}(\mathbb{C}^m); k)$-modules for any field $k$.

7. Virtual Hodge polynomial.

For every algebraic scheme over $\mathbb{C}$ we can define its virtual Hodge polynomial by virtue of Deligne’s mixed Hodge theory ([De1] and [De2]). In this section, we calculate the virtual Hodge polynomials of the algebraic varieties $B_n(m)_B$, $\text{Ch}_n(m)_B$, and $\text{Rep}_n(m)_B$ over $\mathbb{C}$. By these calculations we can determine the virtual Poincaré polynomials of the moduli of absolutely irreducible representations of degree 2 for free monoids (See [Na3]).

7.1. Definition of virtual Hodge polynomial. In this subsection, we give a survey on virtual Hodge polynomials. More precisely, see [DK], [Ch1], [Ch2] and so on.

For an algebraic scheme $X$ over $\mathbb{C}$, we can define the virtual Hodge polynomial $H(X; x, y)$ of $X$ in $\mathbb{Z}[x, y]$ which satisfies the following properties:

1. For a smooth projective variety $X$ over $\mathbb{C}$,

$$H(X; x, y) = \sum_{p, q} h^{p, q}(X) x^p y^q,$$

where $h^{p, q}(X)$ is the $(p, q)$-th Hodge number of $X$.

2. Let $U$ be a Zariski open subset of $X$. Set $Z := X \setminus U$. Then we have

$$H(X; x, y) = H(U; x, y) + H(Z; x, y).$$

3. Let $f : E \to B$ be a fibre bundle with fibre $F$ which has a local trivialization with respect to Zariski topology. Then we have

$$H(E; x, y) = H(B; x, y) H(F; x, y).$$
For a bijective morphism $f : X \to Y$, $H(X; x, y) = H(Y; x, y)$.

**Remark 7.1.** For the virtual Hodge polynomial $H(X; x, y)$ of an algebraic scheme $X$ over $\mathbb{C}$, we call the polynomial $H(X; t, t) \in \mathbb{Z}[t]$ the virtual Poincaré polynomial of $X$. See also [Fu] §4.5 for virtual Poincaré polynomials.

**Example 7.2.** By the above properties we easily obtain

$$H(\mathbb{P}^n; x, y) = 1 + xy + x^2y^2 + \cdots + x^ny^n,$$

$$H(\mathbb{C}^n; x, y) = x^ny^n.$$

**Notation 7.3.** In the sequel, we put $z = xy$.

Let us calculate several virtual Hodge polynomials.

**Example 7.4.** The virtual Hodge polynomial of $GL_n(\mathbb{C})$ can be calculated as follows: The group $GL_n(\mathbb{C})$ acts on $\mathbb{P}^{n-1}$ canonically. The stabilizer of $(1 : 0 : \cdots : 0)$ is isomorphic to $GL_{n-1}(\mathbb{C}) \times GL_1(\mathbb{C}) \times \mathbb{C}^{n-1}$ as an algebraic scheme. By considering the fibre bundle $GL_{n-1}(\mathbb{C}) \times GL_1(\mathbb{C}) \times \mathbb{C}^{n-1} \to GL_n(\mathbb{C}) \to \mathbb{P}^{n-1}$, we have

$$H(GL_n(\mathbb{C})) = H(GL_{n-1}(\mathbb{C}))H(GL_1(\mathbb{C}))H(\mathbb{C}^{n-1})H(\mathbb{P}^{n-1}).$$

Since $H(GL_1(\mathbb{C})) = z - 1$, we obtain

$$H(GL_n(\mathbb{C})) = z^{(n-1)n/2} \prod_{k=1}^{n} (z^k - 1)$$

by induction. We also have

$$H(PGL_n(\mathbb{C})) = z^{(n-1)n/2} \prod_{k=2}^{n} (z^k - 1)$$

by the fibre bundle $GL_1(\mathbb{C}) \to GL_n(\mathbb{C}) \to PGL_n(\mathbb{C})$.

Let $X$ be an algebraic scheme over $\mathbb{C}$. We calculate the virtual Hodge polynomial of the configuration space $F_n(X)$ of $X$. The following proposition has been proved in [FM] Proposition 2.1, essentially.

**Proposition 7.5.** Let $H(X)$ be the virtual Hodge polynomial of $X$. Then the virtual Hodge polynomial $H(F_n(X))$ of $F_n(X)$ is given by

$$H(F_n(X)) = \prod_{k=0}^{n-1} (H(X) - k).$$

**Proof.** We prove the statement by induction on $n$. If $n = 1$, then it is obvious since $F_1(X) = X$. Suppose that the statement is true until $n-1$. The scheme $X \times F_{n-1}(X)$ is a disjoint union of $F_n(X)$ and $(n-1)$ pieces of subschemes
which are isomorphic to \(F_{n-1}(X)\). Hence we have \(H(X)H(F_{n-1}(X)) = H(F_n(X)) + (n - 1)H(F_{n-1}(X))\), which easily follows the statement. \(\square\)

As a corollary, we have:

**Corollary 7.6.** The virtual Hodge polynomial \(H(F_n(C^m))\) of the configuration space \(F_n(C^m)\) is given by

\[
H(F_n(C^m)) = \prod_{k=0}^{n-1} (z^m - k).
\]

### 7.2. The virtual Hodge polynomial of the moduli of representations with Borel mold

In this subsection, we calculate the virtual Hodge polynomials of \(B_n(m)_B\), \(Ch_n(m)_B\), and \(Rep_n(m)_B\).

By Proposition 3.6 we see that \(B_n(m)_B \to F_n(C^m)\) is a fibre bundle with fibre \((C^m \setminus C^1)^{n-1} \times (C^m)^{(n-1)(n-1)}/2\) over \(C\). Hence we have:

**Proposition 7.7.** The virtual Hodge polynomial \(H(B_n(m)_B)\) of \(B_n(m)_B\) is given by

\[
H(B_n(m)_B) = (z^m - z)^{n-1}z^{m(n-2)(n-1)/2} \prod_{k=0}^{n-1} (z^m - k).
\]

Proposition 3.8 follows that \(Ch_n(m)_B \to F_n(C^m)\) is a fibre bundle with fibre \(\mathbb{P}^{m-2} \times (C^{m-1})^{(n-2)(n-1)/2}\) over \(C\). We also see that \(B_n(m)_B \to Ch_n(m)_B\) is a \(B_n(C)\)-principal fibre bundle with respect to Zariski topology. From these facts we have:

**Proposition 7.8.** The virtual Hodge polynomial \(H(Ch_n(m)_B)\) of \(Ch_n(m)_B\) is given by

\[
H(Ch_n(m)_B) = \frac{(z^{m-1} - 1)^{n-1}}{(z - 1)^{n-1}} z^{(m-1)(n-2)(n-1)/2} \prod_{k=0}^{n-1} (z^m - k).
\]

By the fact that \(Rep_n(m)_B \to Ch_n(m)_B\) is a \(\text{PGL}_n(C)\)-principal fibre bundle with respect to Zariski topology, we can calculate \(H(Rep_n(m)_B)\) as follows:

**Proposition 7.9.** The virtual Hodge polynomial \(H(Rep_n(m)_B)\) is given by

\[
H(Rep_n(m)_B) = \frac{(z^m - z)^{n-1}}{(z - 1)^{n-1}} z^{m(n-2)(n-1)/2} \prod_{k=0}^{n-1} (z^m - k) \prod_{k=2}^{n} (z^k - 1).
\]
8. Remarks on the case $m = \infty$.

There is a natural inclusion map $B_n(m)_B \hookrightarrow B_n(m+1)_B$ given by $(A_1, \ldots, A_m) \mapsto (A_1, \ldots, A_m, 0)$. Then we see that this map also induces natural inclusion maps $\text{Ch}_n(m)_B \hookrightarrow \text{Ch}_n(m+1)_B$ and $\text{Rep}_n(m)_B \hookrightarrow \text{Rep}_n(m+1)_B$. We define the spaces $B_n(\infty)_B$, $\text{Ch}_n(\infty)_B$ and $\text{Rep}_n(\infty)_B$ to be the homotopy direct limits (telescopes) of the following systems, respectively:

\[
\begin{align*}
B_n(2)_B & \hookrightarrow B_n(3)_B \hookrightarrow \cdots \hookrightarrow B_n(m)_B \hookrightarrow \cdots \\
\text{Ch}_n(2)_B & \hookrightarrow \text{Ch}_n(3)_B \hookrightarrow \cdots \hookrightarrow \text{Ch}_n(m)_B \hookrightarrow \cdots \\
\text{Rep}_n(2)_B & \hookrightarrow \text{Rep}_n(3)_B \hookrightarrow \cdots \hookrightarrow \text{Rep}_n(m)_B \hookrightarrow \cdots .
\end{align*}
\]

In this section we study $B_n(\infty)_B$, $\text{Ch}_n(\infty)_B$ and $\text{Rep}_n(\infty)_B$.

The inclusion $\mathbb{C}^n \hookrightarrow \mathbb{C}^{n+1}$ given by $(z_1, \ldots, z_m) \mapsto (z_1, \ldots, z_m, 0)$ defines an inclusion $F_n(\mathbb{C}^m) \hookrightarrow F_n(\mathbb{C}^{m+1})$. We denote by $F_n(\mathbb{C}^\infty)$ the homotopy direct limit of the system

\[
F_n(\mathbb{C}^2) \hookrightarrow F_n(\mathbb{C}^3) \hookrightarrow \cdots \hookrightarrow F_n(\mathbb{C}^m) \hookrightarrow \cdots .
\]

The following lemma follows from Lemma 4.2 and the fact that the configuration space $F_n(\mathbb{C}^m)$ is $(2m-2)$-connected.

**Lemma 8.1.** $F_n(\mathbb{C}^\infty)$ and $B_n(\infty)_B$ are weakly contractible.

We recall that there is a fibre bundle

\[
(\mathbb{C}P^{m-2})^{n-1} \longrightarrow \text{Ch}_n(m)_B' \longrightarrow F_n(\mathbb{C}^m).
\]

By the long exact sequence of homotopy groups associated with the fibre bundle, $(\mathbb{C}P^{m-2})^{n-1} \rightarrow \text{Ch}_n(m)_B'$ induces isomorphisms of homotopy groups up to dimension $2m - 3$. There is a commutative diagram

\[
\begin{array}{cccc}
(\mathbb{C}P^{m-2})^{n-1} & \longrightarrow & \text{Ch}_n(m)_B' & \longrightarrow & \text{Ch}_n(m)_B \\
\downarrow & & \downarrow & & \downarrow \\
(\mathbb{C}P^{m-1})^{n-1} & \longrightarrow & \text{Ch}_n(m+1)_B' & \longrightarrow & \text{Ch}_n(m+1)_B, \\
\end{array}
\]

where the vertical arrows are natural inclusions. This diagram induces a map

\[
(\mathbb{C}P^{\infty})^{n-1} \longrightarrow \text{Ch}_n(\infty)_B' \longrightarrow \text{Ch}_n(\infty)_B.
\]

**Proposition 8.2.** $(\mathbb{C}P^{\infty})^{n-1} \rightarrow \text{Ch}_n(\infty)_B$ is a weak homotopy equivalence.

**Proof.** Since $(\mathbb{C}P^{m-2})^{n-1} \rightarrow \text{Ch}_n(m)_B'$ is a homotopy equivalence up to dimension $2m - 3$, the map $(\mathbb{C}P^{\infty})^{n-1} \rightarrow \text{Ch}_n(\infty)_B'$ is a weak homotopy equivalence. The homotopy equivalence $\text{Ch}_n(m)_B' \hookrightarrow \text{Ch}_n(m)_B$ implies that $\text{Ch}_n(\infty)_B' \rightarrow \text{Ch}_n(\infty)_B$ is a weak homotopy equivalence. \qed
The homotopy direct limit of the fibre bundles $T_m \to B_n(m)_B \to \text{Ch}_n(m)_B$ is a model of universal principal $T_\mathbb{R}$-bundle.

**Corollary 8.3.** The cohomology of $\text{Ch}_n(\infty)_B$ is given by

$$H^*(\text{Ch}_n(\infty)_B) \cong \mathbb{Z}[t_1, \ldots, t_{n-1}],$$

where the degree of $t_i$ is 2 for $i = 1, \ldots, n - 1$.

We recall that there is a map from $\text{Rep}_n(m)_B$ to $\text{Flag}(\mathbb{C}^n)$ which is compatible with the inclusions $\text{Rep}_n(m)_B \hookrightarrow \text{Rep}_n(m + 1)_B$. Hence we obtain a map $\text{Rep}_n(\infty)_B \to \text{Flag}(\mathbb{C}^n)$.

**Proposition 8.4.** $\text{Rep}_n(\infty)_B \to \text{Flag}(\mathbb{C}^n)$ is a weak homotopy equivalence. Hence the cohomology ring of $\text{Rep}_n(\infty)_B$ is given by

$$H^*(\text{Rep}_n(\infty)_B) \cong \mathbb{Z}[t_1, \ldots, t_n]/(c_1, \ldots, c_n).$$

**Proof.** The fibre of $\text{Rep}_n(m)_B \to \text{Flag}(\mathbb{C}^n)$ is $B_n(m)_B$. Then the proposition follows from Lemma 8.1. 

**References**


TOPOLOGY OF MODULI OF REPRESENTATIONS


Received May 15, 2002. T. Torii was partially supported by JSPS Research Fellowships for Young Scientists.

DEPARTMENT OF GENERAL EDUCATION
TAKUMA NATIONAL COLLEGE OF TECHNOLOGY
KAGAWA 769–1192
JAPAN
E-mail address: nakamoto@dg.takuma-ct.ac.jp

DEPARTMENT OF APPLIED MATHEMATICS
FUKUOKA UNIVERSITY
FUKUOKA 814-0180
JAPAN
E-mail address: torii@bach.sm.fukuoka-u.ac.jp
THE CONSTANT OF INTERPOLATION

ARTUR NICOLAU, JOAQUIM ORTEGA-CERDÀ, AND KRISTIAN SEIP

We prove that a suitably adjusted version of Peter Jones’ formula for interpolation in $H^\infty$ gives a sharp upper bound for what is known as the constant of interpolation. We show how this leads to precise and computable numerical bounds for this constant.

With each finite or infinite sequence $Z = (z_j) \ (j = 1, 2, \ldots)$ of distinct points $z_j = x_j + iy_j$ in the upper half-plane of the complex plane, we associate a number $M(Z) \in \mathbb{R}^+ \cup \{+\infty\}$ which we call the constant of interpolation. We may define it in two equivalent ways. The first is related to Carleson’s interpolation theorem for $H^\infty$ [Car58]. We say that $Z$ is an interpolating sequence if the interpolation problem

$$f(z_j) = w_j, \ j = 1, 2, \ldots$$

(1)

has a solution $f \in H^\infty$ for each bounded sequence $(w_j)$ of complex numbers. Using the open mapping theorem, we find that if $Z$ is an interpolating sequence, then we can always solve (1) with a function $f$ such that

$$\|f\|_\infty \leq C \| (w_j)\|_\infty$$

for some $C < \infty$ depending only on $Z$. The constant of interpolation $M(Z)$ is declared to be the smallest such $C$. We set $M(Z) = +\infty$ if $Z$ is not an interpolating sequence.

By a classical theorem of Pick (see [Gar81, p. 2]), we may alternatively define $M(Z)$ as follows: Let $M_n(Z)$ be the smallest number $C$ such that the matrices

$$\begin{pmatrix}
1 - \overline{w}_j w_k \\
\overline{z}_j - z_k
\end{pmatrix}_{j,k=1,2,\ldots,n}
$$

are positive semi-definite whenever $\|(w_j)\|_\infty \leq 1/C$. The constant of interpolation is then $M(Z) = M_n(Z)$ if $Z$ is a finite sequence consisting of $n$ points and $M(Z) = \lim_{n\to\infty} M_n(Z)$ if $Z$ is infinite. We will make no use of this definition, but have stated it to make the reader aware of the relevance of $M(Z)$ for the classical Nevanlinna-Pick problem. Note that this connection has been investigated by Koosis, who derived Carleson’s theorem formally from Pick’s theorem [Koo00].
Carleson’s interpolation theorem [Car58] states that $Z$ is an interpolating sequence (or alternatively $M(Z) < \infty$) if and only if
\[ \delta(Z) = \inf_{j \neq k} \prod_{k \neq j} \frac{|z_j - z_k|}{|z_j - z_k|} > 0. \]

Clearly, an interpolating sequence satisfies the Blaschke condition. We let $B$ be the associated Blaschke product and set
\[ B_j(z) = \frac{z - \overline{z_j}}{z - z_j} B(z), \]
so that we may write $\delta(Z) = \inf_j |B_j(z_j)|$.

An interesting result related to Carleson’s theorem is that if $M(Z) < \infty$, then the interpolation may be obtained by means of a linear operator. In fact, P. Beurling [Car63] proved that there exist $f_j \in H^\infty$ with $f_j(z_j) = 1$ and $f_j(z_k) = 0$ if $k \neq j$, such that
\[ M(Z) = \sup_z \sum_j |f_j(z)|. \]

The functions $f_j$ have the form
\[ f_j(z) = \frac{B_j(z)}{B_j(z_j)} \left( \frac{2i y_j}{z - z_j} \right)^2 \frac{G(z)}{G(z_j)}, \]
where $G$ is a bounded analytic function solving a certain nonlinear extremal problem. Unfortunately, $G$ is not given explicitly, and it seems very difficult to get much further. The problem of finding $G$ can be seen as a version of the Nevanlinna-Pick interpolation problem, where one is interested in computing $M(Z)$ and finding solutions of minimal norm. There are classical results of R. Nevanlinna describing these solutions, but they are very implicit and give little help in concrete situations. It is therefore of interest to find more explicit solution operators, along with good estimates for $M(Z)$.

A remarkably simple formula was found by P. Jones [Jon83]. He showed that the series
\[ f(z) = \sum_j w_j \frac{B_j(z)}{B_j(z_j)} \left( \frac{2i y_j}{z - z_j} \right)^2 \exp \left( -a \sum_{y_k \leq y_j} \frac{y_k}{z - z_k} - \frac{y_k}{\overline{z_j} - \overline{z_k}} \right) \]
defines a function $f \in H^\infty$ such that $f(z_j) = w_j$ with $\|f\| \leq C \|(w_j)\|_\infty$. Here $a$ can be chosen freely and $C$ is a constant depending on $a$ and the sequence $Z$. Jones’ formula can be tweaked in several ways. For instance, in [Vin83], Vinogradov found functions $f_j$ that are rational when the interpolating sequence is finite. This is of interest for some applications; see [Nik02, p. 179]. The purpose of this note is to show that this explicit operator, conveniently adjusted, is close to optimal. By considering a certain extreme configuration of points, we are in fact able to prove that it yields
a sharp upper bound for $M(Z)$. As a result, $M(Z)$ may be bounded from above and below by fairly explicit numerical constants.

We begin by showing how to “optimize” Jones’ formula. Take an analytic function $g$ such that $g(i) = 1$. We need $|g|$ to have a harmonic majorant, so we require $(z + i)^{-2}g(z) \in H^1$ [Gar81, p. 60]. Let $u$ denote the least harmonic majorant of $|g|$ and set

$$g_j(z) = g((z - x_j)/y_j), \quad u_j(z) = u((z - x_j)/y_j).$$

We assume further that $g$ is such that

$$U_k(z) = \sum_{y_j \leq y_k} \frac{u_j(z)}{|B_j(z_j)|}$$

defines a harmonic function; let $V_k(z)$ be a harmonic conjugate of $U_k$, and set $G_k = U_k + iV_k$. This leads us to the following interpolation formula:

$$f(z) = \sum_j w_j \frac{B_j(z)}{B_j(z_j)} g_j(z) \exp\left(-a(G_j(z) - G_j(z_j))\right)$$

with $a$ some constant which may be chosen freely. Clearly, $f(z_j) = w_j$. We define

$$c_J(Z, g) = \sup_j U_j(z_j),$$

so that for arbitrary $z$ we get the estimate

$$|f(z)| \leq \|(w_j)\|_\infty \frac{\exp(ac_J(Z, g))}{a} \sum_j \frac{|g_j(z)|}{|B_j(z_j)|} \exp(-aU_j(z)).$$

Replacing $|g_j|$ by $u_j$, we find that the latter sum is a lower Riemann sum for the integral

$$\int_0^\infty e^{-t} dt$$

so that we arrive at the estimate

$$|f(z)| \leq \|(w_j)\|_\infty \frac{\exp(ac_J(Z, g))}{a}.$$

We see that the optimal choice of $a$ is $1/c_J(Z, g)$, and this leads us to the bound

$$M(Z) \leq ec_J(Z, g).$$

We may finally minimize $c_J(Z, g)$ and define

$$c_J(Z) = \inf_g c_J(Z, g)$$

so that

$$M(Z) \leq ec_J(Z).$$
We have then proved one part of the following theorem:

**Theorem 1.** For every sequence $Z$ in the upper half-plane,

$$M(Z) \leq ec_J(Z).$$

The inequality is best possible in the sense that the constant $e$ on the right side of (2) cannot be replaced by any smaller number.

We postpone for the moment the proof of the sharpness of (2); it will be established by means of an explicit example at the end of this note.

It may be argued that finding the $g$ minimizing $c_J(Z,g)$ is not much easier than solving for the function $G$ in P. Beurling’s formula. However, we will now point out that $c_J(Z)$ relates nicely to more computable characteristics.

An immediate observation is that if we choose $g(z) = -4/(z+i)^2$, then

$$c_J(Z,g) = c_H(Z) = \sup_n \sum_{y_j \leq y_n} \frac{4y_j(y_j + y_n)}{|z_j - \bar{z}_n|^2 |B_j(z_j)|^2}.$$ 

This choice of $g$ corresponds to the original version of Jones’ formula. (The letter ‘$H$’ in $c_H(Z)$ stands for Havin; see below.) For this characteristic we have the following result:

**Theorem 2.** For every sequence $Z$ in the upper half-plane,

$$M(Z) \leq kc_H(Z)$$

for some universal constant $k$. The best possible $k$ lies in the interval $[\pi/\log 4, e] = [2.2662\ldots, 2.7183\ldots]$.

We have already established the upper bound for $k$. The lower bound will again follow from the example to be considered below.

Our third and final characteristic was introduced by V. Havin in the first appendix of [Koo98]. Havin’s presentation in [Koo98] was based on work by Vinogradov, Gorin, and Hruščev [VGH81]. We get Havin’s characteristic from the expression for $M(Z)$ obtained from Carleson’s duality argument (see [Gar81, p. 135]):

$$M(Z) = \sup \left\{ 4\pi \sum y_j |h(z_j)| : h \in H^1, \|h\|_1 \leq 1 \right\}.$$

If we choose $h(z) = \pi^{-1} y_k/(z - \bar{z})^2$, $k = 1, 2, \ldots$, we arrive at

$$c_H(Z) = \sup_k \sum_j \frac{4y_ky_j}{|z_k - \bar{z}_j|^2 |B_j(z_j)|}$$

along with the estimate

$$M(Z) \geq c_H(Z).$$
Since clearly \( c_{HJ}(Z) \leq 2c_H(Z) \), we may summarize our findings as a chain of inequalities:

\[
(3) \quad c_H(Z) \leq M(Z) \leq ec_{J}(Z) \leq ec_{HJ}(Z) \leq 2ec_H(Z).
\]

In [Koo98], Havin proves that

\[
c_H(Z) \leq M(Z) \leq kc_H(Z),
\]

with \( k \) a universal constant. To prove the right inequality, he proceeds by duality and uses the invariant Blaschke characterization of Carleson measures, which is closely related to the original proof of Carleson. By computing both \( c_H(Z) \) and \( M(Z) \) when \( Z \) consists of two points, he also shows that the left inequality is best possible. In fact, it may be checked that each of the inequalities in our chain (3) is sharp.

To interpret the “geometric” contents of our characteristics, it may be useful to relate them to the condition

\[
(4) \quad \sup_k \sum_j \frac{d_{j/k}}{|z_k - \bar{z}_j|^2} < +\infty,
\]

which is called the invariant Blaschke condition (see [Gar81, p. 239]). We see that our three characteristics are closely related to the supremum appearing in (4). It may also be noted that by the bound \( M(Z) \leq 2ec_H(Z) \) and a calculus argument applied to the invariant Blaschke sum, we obtain

\[
M(Z) \leq \frac{2e + 4e \log(1/\delta(Z))}{\delta(Z)};
\]

see [Koo98, p. 268].

We finally turn to our example which proves the sharpness of (2) and the lower bound for \( k \) in Theorem 2. In what follows the notation \( a(\gamma) \sim b(\gamma) \) will mean that \( a(\gamma) \) and \( b(\gamma) \) are asymptotically equal, i.e.,

\[
\lim_{\gamma \to +\infty} a(\gamma)/b(\gamma) = 1.
\]

**An example.** Fix \( \gamma > 0 \) and consider the Blaschke product defined by

\[
B(z) = B(\gamma, z) = \prod_{k \leq 0} \frac{z - ie^{k/\gamma}}{z + ie^{k/\gamma}} \prod_{k > 0} \frac{ie^{k/\gamma} - z}{z + ie^{k/\gamma}}.
\]

The signs have been chosen so that \( iB'(i) > 0 \), which ensures convergence of the product. The sequence of zeros \( Z_{\gamma} = (ie^{k/\gamma})_{k \in \mathbb{Z}} \) is clearly an interpolating sequence with \( M(Z_{\gamma}) \) blowing up when \( \gamma \) tends to \(+\infty\). To obtain appropriate estimates for \( B \), we relate it to the function

\[
F(z) = 2e^{-\frac{z^2}{2}} \sin(\pi \gamma \log(-iz)),
\]

where \( \log(z) \) is the principal branch of the logarithm. Both \( B \) and \( F \) are bounded functions, and they have the same zeros. The quotient \( F(z)/B(z) \)
is an outer function with modulus close to 1 when $\gamma$ is large. More precisely, we have

$$\sup_{x \in \mathbb{R} \setminus \{0\}} \left| \frac{\log |F(x)|}{|B(x)|} \right| \sim e^{-\pi^2 \gamma},$$

and therefore the same asymptotic relation holds in the upper half-plane.

The Blaschke product $B$ is highly symmetric. It is real on the imaginary half-axis $i\mathbb{R}^+$ and moreover $B(e^{1/\gamma} z) = -B(z)$. We check that on $i\mathbb{R}^+$ the modulus of $B$ peaks at the points $\{ie^{(k+1/2)/\gamma} : k \in \mathbb{Z}\}$. Again comparing it to $F$, we check that

$$B(ie^{(k+1/2)/\gamma}) = (-1)^k 2e^{-\pi^2 \gamma^2} t_{\gamma} \quad \text{with} \quad t_{\gamma} \sim 1.$$

We will now obtain a lower estimate for $M(Z_{\gamma})$ by finding a minimal norm solution of the interpolation problem

$$f(ie^{k/\gamma}) = (-1)^k, \quad k \in \mathbb{Z}.$$

By (5), the problem is solved by the function

$$g(z) = c_{\gamma} B(e^{1/(2\gamma)} z),$$

with $c_{\gamma}$ an appropriate constant satisfying $c_{\gamma} \sim e^{\pi^2 \gamma^2}/2$. This means that if we can prove that $g$ is a minimal norm solution, then it follows that

$$M(Z_{\gamma}) \geq \frac{t_{\gamma} e^{\pi^2 \gamma^2}}{2} \quad \text{with} \quad t_{\gamma} \sim 1.$$

We wish to prove that $g$ is a solution of minimal norm. To this end, observe that an arbitrary minimal norm solution can expressed as

$$f = g + hB$$

with $h$ a bounded analytic function. We may assume that $f$ is real on $i\mathbb{R}^+$ because by symmetry we may if necessary replace $f$ by $(f(-\overline{z}) + f(z))/2$. Thus $h$ is also real on $i\mathbb{R}^+$. We define

$$h_m(z) = \frac{1}{m} \sum_{k=0}^{m-1} h(e^{2k/\gamma} z),$$

and choose a convergent subsequence $h_{m_k}(z) \to \tilde{h}(z)$ such that the limit function satisfies $\tilde{h}(e^{2/\gamma} z) = \tilde{h}(z)$, and $\tilde{h}(iy) \in \mathbb{R}$ for real $y$. Hence $\tilde{f} = g + \tilde{h}B$ is also a minimal norm solution and $\tilde{f}(e^{2/\gamma} z) = f(z)$. Finally, note that

$$\varphi(z) = \frac{1}{2} (\tilde{f}(z) - \tilde{f}(e^{1/\gamma} z)),$$

is a minimal norm solution as well such that

$$\varphi(e^{1/\gamma} z) = -\varphi(z).$$
Assume now that \( g \) is not a minimal norm solution. Then \( \| \varphi \|_\infty < \| g \|_\infty \). Between the points \( i \) and \( ie^{1/\gamma} \), \( \varphi \) has a zero \( i\delta \) because it is real on \( i\mathbb{R} \). Therefore, by the periodicity expressed by (7), \( \varphi \) has zeros at \( i\delta e^{k/\gamma}, k \in \mathbb{Z} \). It follows that we may factorize \( \varphi \) as
\[
\varphi(z) = B(z/\delta) \varphi_0(z).
\]

We evaluate \( \varphi \) at the point \( i \) and get
\[
1 = |\varphi(i)| = |B(i/\delta)| |\varphi_0(i)| \leq \frac{1}{c_\gamma} \| \varphi \|_\infty = \frac{\| \varphi \|_\infty}{c_\gamma} < 1,
\]
which is a contradiction. We conclude that \( g \) has minimal norm so that (6) holds.

The next step is to compute \( c_f(Z_\gamma) \). Since \( B(e^{1/\gamma} z) = -B(z) \), we have that
\[
|e^{k/\gamma} B'(ie^{k/\gamma})| = |B'(i)|
\]
for each integer \( k \). Hence
\[
|B_k(ie^{k/\gamma})| = 2e^{k/\gamma}|B'(ie^{k/\gamma})| = 2|B'(i)|.
\]
The derivative \( B'(i) \) can be estimated in terms of \( F'(i) \), which gives us
\[
iB'(i)e^{\frac{\pi^2}{2}}/(2\pi\gamma) \to 1 \text{ as } \gamma \to +\infty.
\]

Thus
\[
c_f(Z_\gamma) \sim (4\pi\gamma)^{-1} e^{\frac{\pi^2}{2}} \inf_{g(i)=1} \sup_{k \in \mathbb{Z}} \sum_{y_j \leq y_k} u(iy_k/y_j)
\]
with \( u \) denoting as before the least harmonic majorant of \( |g| \). Using the explicit expression for this majorant, we get
\[
\inf_{g(i)=1} \sup_{k \in \mathbb{Z}} \sum_{y_j \leq y_k} u(iy_k/y_j) = \inf_{g(i)=1} \sum_{k \geq 0} \frac{1}{\pi} \int_{\mathbb{R}} \frac{1}{1 + t^2} |g(e^{k/\gamma} t)| dt.
\]

We interpret the sum on the right as a Riemann sum, so that
\[
\sum_{y_j \leq y_k} u(iy_k/y_j) \sim \frac{\gamma}{\pi} \int_{\mathbb{R}} \frac{1}{1 + t^2} \int_0^\infty |g(te^x)| dx\, dt
\]
\[
= \frac{\gamma}{\pi} \int_{\mathbb{R}} \frac{1}{1 + t^2} \int_t^\infty \frac{|g(u)|}{u} du\, dt.
\]
Integrating by parts, we get
\[
\sum_{y_j \leq y_k} u(iy_k/y_j) \sim \frac{\gamma}{\pi} \int_{\mathbb{R}} \frac{\arctan(t/\gamma)}{t} |g(t)| dt.
\]
We want to minimize the latter integral over all functions $g$ such that $(z + i)^{-1}g \in H^1$ and $g(i) = 1$. This can be restated as an extremal problem in the weighted Hardy space with norm

$$\|h\|^2 = \int_{\mathbb{R}} |h(t)|^2 \frac{\arctan t}{t} \, dt.$$ 

In turn, we can reduce this problem to one for the standard Hardy space $H^2$, and we find that our original problem is solved by the function

$$g_0(z) = \left( \frac{2i}{z + i} \right)^2 \frac{\psi(z)}{\psi(i)},$$

where $\psi(z)$ is the outer function whose modulus is $t/\arctan t$ on $\mathbb{R}$. Since

$$\int_{\mathbb{R}} |g_0(t)| \frac{\arctan t}{t} \, dt = \int_{\mathbb{R}} \frac{4}{(t^2 + 1)} \frac{1}{|\psi(i)|} = \frac{4\pi}{|\psi(i)|},$$

we get

(9) $$c_J(Z_\gamma) \sim \frac{e^{\frac{x_\gamma^2}{2}}}{\pi |\psi(i)|}$$

when plugging our extremal function $g_0$ into (8).

We are left with the computation of $|\psi(i)|$. We first note that

$$\psi(i) = i \exp\left(-\frac{i}{\pi} \int_{\mathbb{R}} \left( \frac{1}{i - t} + \frac{t}{t^2 + 1} \right) \log |\arctan(t)| \, dt \right).$$

Since

$$|\arctan(t)| = \frac{1}{2} \left| \log \left( \frac{1 - it}{1 + it} \right) \right|,$$

the change of variables

$$e^{i\theta} = \frac{1 - it}{1 + it}$$

brings us to the explicit expression

$$\psi(i) = 2i \exp\left(-\frac{1}{2\pi} \int_{-\pi}^{\pi} \log |\log e^{i\theta}| \, d\theta \right) = \frac{2ie}{\pi}.$$

Combining (6) and (9), we conclude that

$$c_J(Z_\gamma) \sim \frac{1}{2e} e^{\frac{x_\gamma^2}{2}} \leq \frac{t_\gamma}{e} M(Z_\gamma) \quad \text{with} \quad t_\gamma \sim 1,$$

which proves the sharpness of (2) of Theorem 1.

The computation of $c_{HJ}(Z_\gamma)$ is straightforward. Indeed,

$$c_{HJ}(Z_\gamma) \sim (4\pi)^{-1} e^{\frac{x_\gamma^2}{2}} \sum_{k \geq 0} \frac{4e^{-k/\gamma}}{(1 + e^{-k/\gamma})}.$$
The sum is again regarded as a Riemann sum, i.e.,

\[ \sum_{k \geq 0} \frac{e^{-k/\gamma}}{1 + e^{-k/\gamma}} \sim \gamma \int_0^\infty \frac{e^{-x}}{1 + e^{-x}} \, dx = \gamma \log 2 \]

so that we arrive at the relation

\[ c_{HJ}(Z_\gamma) \sim \frac{\log 2}{\pi} e^{\frac{x_\gamma}{2}} \leq \frac{t \gamma 2 \log 2}{\pi} M(Z_\gamma) \text{ with } t \gamma \sim 1. \]

This proves the lower bound for \( k \) in Theorem 2.

**Acknowledgments.** We are grateful to N. Nikolski for communicating to us several relevant remarks on the bibliography of this paper.

**References**


Received January 27, 2003. The authors are supported by the European Commission Research Training Network HPRN-CT-2000-00116. The first two authors are supported by DGICYT grants: BFM2002-00571, BFM2002-04072-C02-01 and by the CIRIT:
2001SGR00172, 2001SGR00431. The third author is partly supported by a grant from the Research Council of Norway.

Dept. Matemàtiques
Universitat Autònoma de Barcelona
08193 Bellaterra
Spain
E-mail address: artur@mat.uab.es

Dept. Matemàtica Aplicada i Anàlisi
Universitat de Barcelona
Gran Via 585
08071 Barcelona
Spain
E-mail address: quim@mat.ub.es

Dept. of Mathematical Sciences
Norwegian University of Science and Technology
N–7491 Trondheim
Norway
E-mail address: seip@math.ntnu.no
CONTENTS

Volume 213, no. 1 and no. 2

C. Adams, A. Colestock, J. Fowler, W. Gillam and E. Katerman: Cleanliness of geodesics in hyperbolic 3-manifolds 201

Nils Byrial Andersen: Real Paley-Wiener theorems for the inverse Fourier transform on a Riemannian symmetric space 1

Frank W. Anderson and Barbara K. D’Ambrosia: Rings with sovable quivers 213

Ulrich Clarenz and Heiko von der Mosel: On surfaces of prescribed F-mean curvature 15

A. Colestock with C. Adams, J. Fowler, W. Gillam and E. Katerman 201

Antonio F. Costa and Milagros Izquierdo: Symmetries of real cyclic p-gonal Riemann surfaces 231

Mauro Costantini and Giovanni Zacher: The finite simple groups have complemented subgroup lattices 245

Barbara K. D’Ambrosia with Frank W. Anderson 213

Chongying Dong and Geoffrey Mason: Holomorphic vertex operator algebras of small central charge 253

Robert Finn with Danzhu Shi 111

J. Fowler with C. Adams, A. Colestock, W. Gillam and E. Katerman 201

W. Gillam with C. Adams, A. Colestock, J. Fowler and E. Katerman 201

M. Hickel: Calcul de la fonction d’Artin–Greenberg d’une branche plane 37

Milagros Izquierdo with Antonio F. Costa 231

Frazer Jarvis: Correspondences on Shimura curves and Mazur’s principle at p 267

E. Katerman with C. Adams, A. Colestock, J. Fowler and W. Gillam 201

Seick Kim: Harnack inequality for nondivergent elliptic operators on Riemannian manifolds 281

N.P. Kishye: The integral kernel in the Kuznetsov sum formula for SU(n + 1, 1) (II). The case of one dimensional K-types 295

Weigu Li, Jaume Llibre and Xiang Zhang: Melnikov functions for period annulus, nondegenerate centers, heteroclinic and homoclinic cycles 49

Jaume Llibre with Weigu Li and Xiang Zhang 49
Guangcun Lu: An explicit isomorphism between Floer homology and quantum homology 319

Geoffrey Mason with Chongying Dong 253

Emilia Mezzetti and Orsola Tommasi: Some remarks on varieties with degenerate Gauss image 79

Kazunori Nakamoto and Takeshi Torii: Topology of the moduli of representations with Borel mold 365

Artur Nicolau, Joaquim Ortega-Cerdà and Kristian Seip: The constant of interpolation 389

Joaquim Ortega-Cerdà with Artur Nicolau and Kristian Seip 389

Kristian Seip with Artur Nicolau and Joaquim Ortega-Cerdà 389

Danzhu Shi and Robert Finn: On a theorem of Lancaster and Siegel 111

Orsola Tommasi with Emilia Mezzetti 79

Takeshi Torii with Kazunori Nakamoto 365

Constantin Vernicos: The macroscopic sound of tori 121

Heiko von der Mosel with Ulrich Clarenz 15

C.M. Weinbaum: Word Equation $ABC = CDA$, $B \neq D$ 157

Giovanni Zacher with Mauro Costantini 245

Xiang Zhang with Weigu Li and Jaume Llibre 49

Xiang Zhang with Weigu Li and Jaume Llibre 49
Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use \LaTeX, but papers in other varieties of \TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as \LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of Bib\TeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal’s preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.
Cleanliness of geodesics in hyperbolic 3-manifolds
C. Adams, A. Colesstock, J. Fowler, W. Gillam and E. Katerman
201

Rings with solvable quivers
Frank W. Anderson and Barbara K. D’Ambrosia
213

Symmetries of real cyclic \( p \)-gonal Riemann surfaces
Antonio F. Costa and Milagros Izquierdo
231

The finite simple groups have complemented subgroup lattices
Mauro Costantini and Giovanni Zacher
245

Holomorphic vertex operator algebras of small central charge
Chongying Dong and Geoffrey Mason
253

Correspondences on Shimura curves and Mazur’s principle at \( p \)
Frazer Jarvis
267

Harnack inequality for nondivergent elliptic operators on Riemannian manifolds
Seick Kim
281

The integral kernel in the Kuznetsov sum formula for \( SU(n+1,1) \) (II). The case of one dimensional \( K \)-types
N.P. Kissye
295

An explicit isomorphism between Floer homology and quantum homology
Guangcun Lu
319

Topology of the moduli of representations with Borel mold
Kazunori Nakamoto and Takeshi Torii
365

The constant of interpolation
Artur Nicolau, Joaquim Ortega-Cerdà and Kristian Seip
389