



# On the classification of rank-two representations of quasiprojective fundamental groups

Kevin Corlette and Carlos Simpson

## ABSTRACT

Suppose that  $X$  is a smooth quasiprojective variety over  $\mathbb{C}$  and  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, \mathbb{C})$  is a Zariski-dense representation with quasiunipotent monodromy at infinity. Then  $\rho$  factors through a map  $X \rightarrow Y$  with  $Y$  either a Deligne–Mumford (DM) curve or a Shimura modular stack.

## 1. Introduction

Let  $X$  be a connected smooth quasiprojective variety over  $\mathbb{C}$  with basepoint  $x$ . We look at representations  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, \mathbb{C})$ . We assume throughout that *the monodromy at infinity is quasiunipotent*. If  $X \subset \overline{X}$  is a normal-crossings compactification with complementary divisor  $D = \sum D_i$ , and if  $\gamma_i$  are loops going around the components  $D_i$ , this condition means that the  $\rho(\gamma_i)$  are quasiunipotent, in other words their eigenvalues are roots of unity.

A representation  $\rho$  is *Zariski-dense* if the Zariski-closure of its image is the whole group  $\mathrm{SL}(2, \mathbb{C})$ . A reductive representation of rank two is either Zariski-dense or otherwise it becomes reducible upon pullback to a finite unramified covering of  $X$ . We classify representations  $\rho$  which are Zariski-dense and have quasiunipotent monodromy at infinity. See [Ara97, Bea92, Del07, Dim06, Fy07] for a similar classification in the reducible case.

The geometry of the fundamental group of an algebraic variety has been studied from many different angles [ACT02, ABCKT96, Ara97, ABR92, Cam94, Cat96, Don85, Gro89, JY86, KM98, Kol93, Pan04, Rez02, RT97, Tol90, Vie05]. The methods we use here are based on the theory of harmonic mappings, both to symmetric spaces and combinatorial complexes [Cor88, Cor91, Del06, Don87, Eys04, GS92, JZ96, JZ00, Kat97, KR98, Kli03, Moc07, NR08, Sim92, Sim91, Siu80, Zuo94].

Our classification is obtained by looking at the interplay between different properties of  $\rho$ . The main property is *factorization*: we say that  $\rho$  *factors through a map*  $f : X \rightarrow Y$  if it is isomorphic to the pullback of a representation of  $\pi_1(Y, f(x))$ . This notion can be extended in two ways, for example,  $\rho$  *projectively factors* through  $f$  if the projected representation into  $\mathrm{PSL}(2, \mathbb{C})$  factors through  $f$ . The other extension is that it is convenient—and basically almost essential—to look at the notion of factorization through maps to *Deligne–Mumford (DM) stacks*  $Y$  rather than just varieties. In a certain sense this takes the place of complicated statements involving coverings of  $X$ . It even subsumes the notion of projective factorization, because projective factorization is equivalent to factorization through a new DM-stack obtained by putting a stack structure with group  $\mathbb{Z}/2$  (the center of  $\mathrm{SL}(2, \mathbb{C})$ ) over the generic point of  $Y$ .

One of the main cases of factorization we are concerned with is factorization through a curve. A smooth one-dimensional DM-stack is called a *DM-curve*. Recall that an *orbicurve* is a DM-curve whose generic stabilizer is trivial. An orbicurve is given by the data of a smooth curve together

---

Received 27 February 2007, accepted in final form 20 February 2008.

*2000 Mathematics Subject Classification* 58E20 (primary), 14D07, 22E40, 32J27 (secondary).

*Keywords*: fundamental group, representation, harmonic map, tree, Deligne–Mumford stack, Shimura variety.

This journal is © [Foundation Compositio Mathematica](#) 2008.

with a collection of marked points that are assigned integer ( $\geq 2$ ) weights. Factorization through a DM-curve is equivalent to projective factorization through an orbicurve (Corollary 3.3).

The other case we need to consider arises when the representation is motivic; in fact, it comes from a family of abelian varieties. The families of abelian varieties whose monodromy representations break up into rank-two pieces are given by maps to certain Shimura varieties or stacks. These Shimura varieties are closely analogous to Hilbert modular varieties. However, Hilbert modular varieties parametrize abelian varieties with real multiplication, while in general we need to look at abelian varieties with multiplication by a totally imaginary extension of a totally real field. The condition that the tautological representation goes into  $\mathrm{SL}(2)$  basically says that the universal covering of the Shimura variety is a polydisk, that is, a product of one-dimensional disks. We work without level structure and call these *polydisk Shimura modular DM-stacks*. A classical example is the case of Shimura curves. The precise construction is reviewed in §9. If  $H$  is a polydisk Shimura DM-stack then  $\pi_1(H)$  has a tautological representation into  $\mathrm{SL}(2, L)$  for a totally imaginary extension  $L$  of a totally real field, and this gives a collection of tautological representations into  $\mathrm{SL}(2, \mathbb{C})$  indexed by the embeddings  $\sigma : L \rightarrow \mathbb{C}$ .

Classifying our rank-two Zariski-dense representations that are quasiunipotent at infinity will consist then of showing that any such representation factors through a map  $f : X \rightarrow Y$ , with  $Y$  being either a DM-curve or a polydisk Shimura modular DM-stack. We consider as ‘known’ the representations on these target stacks  $Y$ . There may be some overlap between these two cases, but one of our basic tasks is to have properties which determine which case of the classification we want to prove for a given representation.

Since we are looking at representations on quasiprojective varieties, we define *rigidity* in a way that takes into account the monodromy at infinity. Fix a normal crossings compactification of  $X$ . For each component  $D_i$  of the divisor at infinity, we have a well-defined conjugacy class of elements of  $\pi_1(X, x)$  corresponding to a loop  $\gamma_i$  going around that component. Thus, for a given representation  $\rho$  this gives a conjugacy class  $C_i$  in the target group. We are assuming that these monodromy elements are quasiunipotent, so  $C_i$  is a quasiunipotent conjugacy class. We can define an affine variety  $R(X, x, \mathrm{SL}(2), \{\overline{C}_i\})$  of representations such that the monodromies  $\rho(\gamma_i)$  are contained in the closures of the  $C_i$ . Let  $M(X, \mathrm{SL}(2), \{\overline{C}_i\})$  denote its universal categorical quotient by the conjugation action. We say that  $\rho$  is *rigid* if it represents an isolated point in the moduli space  $M(X, \mathrm{SL}(2), \{\overline{C}_i\})$  obtained by looking at its own conjugacy classes. In the case of a Zariski-dense representation, this form of rigidity means that there is no non-isotrivial family of representations all having the same conjugacy classes at infinity, going through  $\rho$  (Lemma 6.5).

A property that plays a similar role but that is easier to state is *integrality*. Say that a representation  $\rho$  is *integral* if it is conjugate, in  $\mathrm{SL}(2, \mathbb{C})$ , to a representation  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, A)$  for  $A$  a ring of algebraic integers. For Zariski-dense representations, this is equivalent to asking that the traces  $\mathrm{Tr}(\rho(\gamma))$  be algebraic integers for all  $\gamma \in \pi_1(X, x)$ .

Say that  $\rho$  *comes from a complex variation of Hodge structure* if there is a structure of complex variation of Hodge structure on the corresponding local system  $V$ .

The main relationship between all of these notions is the following first result.

**THEOREM 1.** *Suppose that  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, \mathbb{C})$  is a representation with quasiunipotent monodromy at infinity, such that  $\rho$  does not projectively factor through an orbicurve or, equivalently,  $\rho$  does not factor through a map to a DM-curve. Then  $\rho$  is rigid and integral. Rigidity implies that  $\rho$  comes from a complex variation of Hodge structure.*

This is already known in the case when  $X$  is projective from [Sim91] for rigidity, and for integrality [GS92] and [Sim93], the latter of which was designed to support the original dormant version of this paper. The variation of Hodge structure follows from [Cor88]; see also [Sim92]. In the present,

we extend the result to the quasiprojective case for representations with quasiunipotent monodromy at infinity. The various statements in Theorem 1 appear as Theorems 6.8, 7.3 and 8.1 below.

The underlying argument for both rigidity and integrality comes from Theorem 5.13 about harmonic maps to Bruhat–Tits trees. This strategy is perhaps worth commenting on. It has its origins in the work of Bass [Bas80], Serre [Ser80, Ser77], Culler–Shalen [CS83] and Gromov–Schoen [GS92].

It would certainly have been possible to treat the rigidity question using harmonic maps to symmetric spaces [Cor88, DO85, Don87, ES64, Moc07]. For integrality, though, it is necessary to use the theory of harmonic maps to Bruhat–Tits trees [GS92]. Furthermore, there is a sort of analogy between the two notions: integrality means that a representation into  $\mathrm{SL}(2, \mathbb{Q}_p)$  goes into a compact subgroup, whereas rigidity may be thought of as saying that a representation into  $\mathrm{SL}(2, \mathbb{C}(t))$  goes into a compact subgroup, much as in [CS83]. So, we thought it would be interesting to use harmonic maps to trees to treat both cases at once. This reduces the volume of material about harmonic maps. On the other hand it introduces an additional difficulty, because the Bruhat–Tits tree for  $\mathrm{SL}(2, \mathbb{C}(t))$  is not locally compact. We overcome this by making a reduction to the case of representations in  $\mathrm{SL}(2, \mathbb{F}_q(t))$ , where  $\mathbb{F}_q$  is a finite field. This reduction is fairly standard but it requires a finiteness theorem for the number of possible maps to a hyperbolic DM-curve (Proposition 2.8), see Theorem 6.8. Delzant has independently found this proof of the fact that non-factorization implies rigidity. Some pieces of his proof, including the finiteness statement, appear in [Del07], but that paper is oriented towards proving a similar statement for representations in a solvable group. The classification for solvable representations in [Bea92] and [Del07] complements the present paper because we restrict here to Zariski-dense representations, which are irreducible over any finite covering.

Suppose now that  $\rho$  does not factor through a curve. From the above results we obtain that  $\rho$  is rigid, integral, and comes from a complex variation of Hodge structure. Let  $A$  be a ring of algebraic integers such that  $\rho$  is defined over  $A$ . Let  $\sigma : A \rightarrow \mathbb{C}$  be an embedding (not necessarily the identity one). Let  $\rho^\sigma$  be the composed representation

$$\pi_1(X, x) \xrightarrow{\rho^A} \mathrm{SL}(2, A) \xrightarrow{\sigma} \mathrm{SL}(2, \mathbb{C}).$$

Then  $\rho^\sigma$  is rigid too. Hence,  $\rho^\sigma$  comes from a complex variation of Hodge structure for every  $\sigma$ . We use this data to construct a factorization of  $\rho$  through a polydisk Shimura modular DM-stack  $H$ . As before this means that there is a map  $f : X \rightarrow H$  such that  $\rho$  is the pullback  $f^*$  of one of the tautological rank-two local systems on  $H$ . The main idea is that because the rank is two, the Hodge types can be chosen to be  $(1, 0)$  and  $(0, 1)$ ; then, by integrality, we obtain a family of abelian varieties. The construction of the map to  $H$  is straightforward but it has to take into account the notion of polarization and the special structure of abelian varieties whose Hodge structures split into rank-two pieces over a totally imaginary field. This gives the statement of our main classification result, see § 11.

**THEOREM 2.** *Suppose that  $X$  is a smooth quasiprojective variety and  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, \mathbb{C})$  is a Zariski-dense representation. Suppose that the monodromy transformations around components of the divisor at infinity are quasiunipotent. Then either  $\rho$  comes from a map  $f : X \rightarrow Y$  to a DM-curve  $Y$  or otherwise  $\rho$  comes from pullback of one of the tautological representations by a map  $f : X \rightarrow H$  to a polydisk Shimura modular DM-stack  $H$ .*

The two cases described in this theorem can overlap: there can be rigid local systems on an orbicurve. For any rank such things are classified by Katz’s algorithm [Kat96], but in the rank-two case they can be seen explicitly as hypergeometric systems (see § 7.1).

This paper is a project which we have been entertaining since around 1990. It was motivated by Gromov’s paper [Gro89], and spurred on by a lecture by Schoen in Chicago about [GS92]. In the

projective case the paper [Sim93] was done in the context of this program, in order to obtain the proof of Theorem 1 and thus Theorem 2; in particular, [Sim93] should be considered as an integral part of the present project.

An important element was Larsen's explanation of how to obtain information on the field of definition of the representation. His argument, first reported in [Sim92, Lemma 4.8], plays a crucial role in starting off §10 below.

As we have taken such a long time to write up the classification result, the ambient state of technology has evolved in the meantime [Biq96, JZ00, Kat94, Moc07, Pan04] which makes it reasonable to give statements for quasiprojective varieties. Another recent advance is that the notion of a DM-stack has become standard and well understood, for example Behrend–Noohi [BN06] have classified DM-curves in a way which is very useful for our considerations.

One of the reasons for getting back to this project is that recently there have been some explicit constructions of rank-two local systems on quasiprojective varieties, which in some cases can be seen as coming from projective varieties by passing to a finite ramified cover. The examples we know of are those of Boalch [Boa05] and Panov [Pan04]. It would be interesting to determine the factorizations for these examples explicitly, but we do not treat that question here. That question seems to be answered in some cases by a recent paper of Ben Hamed and Gavrilov [BG05] which gives an explicit geometric origin for solutions of Painlevé VI equations. Kontsevich has a number of conjectures about local systems on curves [Kon07].

## 2. Local systems on DM stacks

In order to obtain optimal statements of the type we consider throughout this paper, it is convenient to consider the notion of a local system over a DM stack.

Recall that a DM *stack* is a 1-stack  $X$  on the site of schemes over  $\mathbb{C}$  with the étale topology, such that there exists a surjective morphism of stacks  $f : Z \rightarrow X$  from a scheme  $Z$  to  $X$ , with  $f$  'representable and étale'. These conditions mean that if  $Y$  is any scheme, then  $Z \times_X Y \rightarrow Y$  is an étale morphism of schemes. We refer the reader to [AV02, BN06, Cam04, DM69, LM00, Noo04] for general references about this notion. In practice, in the cases we consider in this paper,  $X$  will often be the stack theoretical quotient of a smooth variety  $Z$  by an action of a finite group  $G$ . This is the case when  $X$  is a Shimura modular stack, and almost always the case when  $X$  is an orbicurve or DM-curve. In the Shimura modular case,  $X$  has many Galois coverings by Shimura modular varieties obtained by imposing some level structure. In the one-dimensional case, aside from a small number of degenerate situations, an orbicurve has a Galois covering which is a smooth curve (see Lemmas 2.3 and 2.6).

If  $X$  is a DM-stack, then we obtain its *coarse moduli space*  $X^{\text{coarse}}$  which is the universal algebraic space with a map from  $X$ . This exists, and the map  $X \rightarrow X^{\text{coarse}}$  is finite, by Keel–Mori [KM97].

**THEOREM 2.1.** *If  $X$  is a DM-stack, there is a Zariski open covering of  $X^{\text{coarse}}$  such that the pullbacks of the open sets are quotient stacks by finite group actions.*

*Proof.* This is stated in [Toe99, Proposition 1.17], which gives a reference to [Vis89, 2.8] as the proof, which should be taken in light of [KM97]. A sketch of proof is given in [AV02, Lemma 2.2.3.]  $\square$

Owing to this theorem, the reader may without danger imagine that the words 'DM-stack', and so forth, basically mean varieties modulo finite group actions. The *irreducible components* of a DM-stack will by definition be the irreducible components of the coarse moduli space, which pull back to closed substacks of  $X$ .

Local properties of  $X$  are defined by requiring the same properties for the étale covering scheme  $Z \rightarrow X$  occurring in the above definition. In particular, a DM-stack  $X$  is *smooth* if for one (or, equivalently, any) surjective étale map from a scheme  $Z \rightarrow X$ , the scheme  $Z$  is smooth. We restrict our attention in this paper to smooth DM-stacks. A *point* is a morphism  $x : \text{Spec}(\mathbb{C}) \rightarrow X$ , which we denote (by abusive of notation) by  $x \in X$ . A point can have ‘automorphisms’, which is the phenomenon new to stacks.

The notion of a DM-stack is a generalization and transfer to the algebraic category, of the classical notion of ‘orbifold’ or  $V$ -manifold [Kaw78, Sat56, Sat57]. This notion adapts to the algebraic category: an *orbifold* is a smooth DM-stack such that the general point of any irreducible component has trivial automorphism group. The associated complex analytic stack in this case is exactly a complex analytic  $V$ -manifold. Then there is a dense Zariski-open subset of the coarse moduli space  $X^{\text{coarse}}$  over which the projection is an isomorphism, thus an orbifold has a dense open substack which is an algebraic space.

If  $X = Z/G$  is a quotient stack by the action of a finite group (with the quotient being irreducible, say), then  $X$  is an orbifold if and only if  $G$  acts faithfully. The coarse moduli space is the classical quotient space of the action. Orthogonal to this case is the important example of the quotient of a single point by a trivial action of a finite group  $G$ : this is a DM-stack denoted by  $BG$ . All of its points are isomorphic and they all have automorphism group isomorphic to  $G$ .

These two examples permit us to form all possible smooth DM-stacks, as shown by the following structure result.

PROPOSITION 2.2. *Suppose that  $X$  is a smooth DM-stack. There is a universal orbifold  $X^{\text{orb}}$  (that is, a smooth DM-stack with trivial generic stabilizers) with a map  $\phi : X \rightarrow X^{\text{orb}}$  such that locally in the étale topology over the base, the map  $\phi$  is isomorphic to the projection of a product of the base orbifold with a  $BG$  (this is usually called a gerb). The coarse moduli spaces of  $X$  and  $X^{\text{orb}}$  coincide.*

For a proof see Behrend and Noohi [BN06].

### 2.1 Fundamental group and local systems

If  $X$  is a DM-stack and  $x \in X$  is a point, then we obtain the *fundamental group*  $\pi_1(X, x)$ . See Noohi [Noo04] for a general discussion. In the quotient case  $X = Z/G$ , the fundamental group may be viewed more simply as an extension

$$1 \rightarrow \pi_1(Z, z) \rightarrow \pi_1(X, x) \rightarrow G \rightarrow 1.$$

The group in the middle may be defined as the set of paths in  $Z$  starting at  $z$  and going to any preimage of  $x$ . Paths are composed by first translating by an appropriate element of  $g$ , then juxtaposing paths.

We also have the notion of a *local system* over  $X$ , which is by now a standard notion, see [LO06] for example, or [Toe99] for the case of  $\mathcal{D}$ -modules. The case of local systems on the moduli stack of hyperelliptic curves was mentioned in a talk by Hain, see [Hai06, p. 12]. A local system is a collection  $L_{Z,f}$  of local systems over schemes  $Z$  for every section  $f : Z \rightarrow X$ , together with functoriality maps: if

$$f : Z \rightarrow X, \quad f' : Z' \rightarrow X$$

are two maps, and if  $g : Z' \rightarrow Z$  is a map together with a natural transformation  $\eta$  from  $f \circ g$  to  $f'$ , then we obtain a map from  $g^*(L_{Z,f})$  to  $L_{Z',f'}$ . These satisfy some natural axioms. We obtain a tensor category of local systems over  $X$ .

If  $L$  is a local system of rank  $r$  over  $X$  and if  $x \in X$  is a basepoint, then the fiber  $L_x$  is a  $\mathbb{C}$ -vector space of dimension  $r$ , and the monodromy is an action of  $\pi_1(X, x)$  on it. If  $X$  is smooth

and irreducible with fixed basepoint  $x$ , then the category of local systems over  $X$  is equivalent to the category of representations of  $\pi_1(X, x)$ .

If  $X = Z/G$  is a quotient stack of a variety  $Z$ , a local system on  $X$  may be seen as a  $G$ -equivariant local system on  $Z$ , that is, a pair  $(V_Z, \alpha)$  where  $V_Z$  is a local system on  $Z$  and  $\alpha$  is an action associating to each  $g \in G$  an isomorphism  $\alpha(g) : g^*V_Z \cong V_Z$ . The action is required to satisfy a cocycle condition: if  $g, h \in G$ , then  $\alpha(g) \circ g^*\alpha(h) = \alpha(gh)$ .

The notion of a local system on a DM-stack beams back down to the world of varieties in the following way: if  $Y$  is a variety and  $f : Y \rightarrow X$  is a map to a DM-stack  $X$ , then for any local system  $V$  on  $X$  the pullback  $f^*V$  is a local system on  $Y$ .

Something new here is that a map  $f$  can have *automorphisms*. Roughly speaking an automorphism of  $f$  is a section of the sheaf of stabilizer groups pulled back over  $f$ . More generally we should speak of *isomorphisms* between maps  $f, g : Y \rightarrow X$ . We have the following functoriality: if  $a : f \Rightarrow g$  is an isomorphism from  $f$  to  $g$ , then we obtain an isomorphism of pullback local systems  $a^*V : f^*V \cong g^*V$ . This satisfies some usual functoriality and associativity identities. This phenomenon will appear in our statements when we want to say that a map is unique.

### 2.2 DM-curves and orbicurves

We are particularly interested in the case of objects of dimension one. A *DM-curve* is a smooth DM-stack of dimension one. An *orbicurve* is an orbifold of dimension one, so an orbifold is a DM-curve with trivial generic stabilizer. In this case Proposition 2.2 says that a DM-curve is always a gerb over a canonical orbicurve. If  $X$  is an orbicurve, then its coarse moduli space  $X^{\text{coarse}}$  is again a smooth curve (this is special to the case of orbicurves: in general the coarse moduli space of a smooth orbifold will have finite quotient singularities).

An étale covering is a finite étale map; over the complex numbers this is the same thing as a finite topological covering space. For DM-curves, étaleness is measured in terms of a local chart for the stack. In practical terms this comes down to saying that the ramification index, the orbifold indices taking correctly into account, is one.

The data of an orbicurve is determined by the smooth curve  $X^{\text{coarse}}$ , together with a finite set of points  $P_j$  and an integer  $n_j \geq 2$  attached to each point. A non-constant morphism from a connected smooth curve to  $X$  is any non-constant map  $f : Z \rightarrow X^{\text{coarse}}$  such that for any  $z \in Z$  lying over some  $P_j$  the ramification order of  $f$  at  $z$  is divisible by  $n_j$ . The map is étale when the ramification at  $P_j$  is exactly  $n_j$ .

We denote an orbicurve by the notation  $(Y, n_1, \dots, n_k)$  with  $Y = X^{\text{coarse}}$  a smooth curve and  $n_i$  the sequence of integers organized in decreasing order (we leave the choice of points  $P_1, \dots, P_k \in X$  out of this notation).

The paper of Behrend and Noohi [BN06] is a very complete description of the possibilities for DM-curves and orbicurves, and the reader is referred there. Recall here the general outlines of their classification which generalizes the classification of curves. An orbicurve is *spherical*, *elliptic*, or *hyperbolic*. The spherical orbicurves are  $\mathbb{P}^1$ , the *drops*  $(\mathbb{P}^1, a)$ , and *footballs*  $(\mathbb{P}^1, a, b)$  as well as the finite list of cases

$$(\mathbb{P}^1, 2, 2, 2), \quad (\mathbb{P}^1, 2, 3, 3), \quad (\mathbb{P}^1, 2, 3, 4), \quad (\mathbb{P}^1, 2, 3, 5).$$

The universal coverings of these orbicurves are  $\mathbb{P}^1$ , drops, or footballs with relatively prime indices.

**LEMMA 2.3.** *An orbicurve that is not spherical as listed above has an étale covering which is a regular curve different from  $\mathbb{P}^1$ . In particular, the associated complex-analytic orbicurve has a contractible universal covering.*

See [BN06], for a proof.

Now we can define the *elliptic* orbicurves to be those which have a covering by either  $\mathbb{A}^1$  or  $\mathbb{G}_m$ , or an elliptic curve (thus, they are those whose universal covering is  $\mathbb{C}$ ); and the *hyperbolic* orbicurves are the remaining ones—those whose universal covering is a disk.

Note also that an elliptic or hyperbolic orbicurve has infinite fundamental group except for the elliptic case where the covering is  $\mathbb{A}^1$ . In all elliptic and hyperbolic cases, the homotopy type is a  $K(\pi_1, 1)$ .

We say that a DM-curve  $X$  is spherical, elliptic, or hyperbolic according to the type of the associated orbicurve  $X^{\text{orb}}$ .

LEMMA 2.4. *Suppose that  $X$  is a DM-curve and  $\rho : \pi_1(X, x) \rightarrow \text{SL}(2, K)$  is a Zariski-dense representation for any infinite field  $K$ . Then  $X$  is hyperbolic.*

*Proof.* Spherical orbicurves have finite fundamental groups, and elliptic orbicurves have fundamental groups which are virtually abelian. Thus, the same holds for spherical and elliptic DM-curves. None of these types of groups can have Zariski-dense representations to  $\text{SL}(2, K)$  for an infinite field  $K$ .  $\square$

In view of this lemma, for factorization questions we will mainly be looking at hyperbolic orbicurves.

LEMMA 2.5. *Suppose that  $X$  is a DM-curve with basepoint  $x$ . Suppose that the underlying orbifold  $Y := X^{\text{orb}}$  is either elliptic or hyperbolic. Then  $\pi_2(Y) = 0$  and the gerb  $p : X \rightarrow Y$  (locally a product with  $BG$ ) is completely determined by the induced extension*

$$1 \rightarrow G \rightarrow \pi_1(X, x) \rightarrow \pi_1(Y, p(x)) \rightarrow 1.$$

*Thus, we may think of  $X$  as being given by the data of an orbicurve  $Y$  plus an extension of the fundamental group by a group  $G$ . If  $Z$  is a scheme, then a submersive map from  $Z$  to  $X$  is the same thing as a submersive map from  $Z$  to  $Y$  plus a lifting of the map on fundamental groups from  $\pi_1(Z, z)$  to  $\pi_1(X, x)$ .*

*Proof.* This is basically the same as Behrend–Noohi [BN06, Proposition 4.7].

If  $X$  is already an orbicurve, then the generic stabilizer group  $G$  is trivial,  $BG = \text{Spec}(\mathbb{C})$  and  $Y = X$ .

In general, let  $Z$  be the coarse moduli space for  $X$ . It is a normal (hence, smooth) curve. Let  $Y$  be the orbicurve obtained by setting, for each point  $z \in Z$ , the orbifold structure at  $z$  to be the ramification degree of the map  $X \rightarrow Z$  over  $z$  (this is different from one for only a finite number of points). There is a unique factorization to a map  $p : X \rightarrow Y$  (note that there is no question of natural transformations here because  $p$  is submersive and the generic stabilizers on  $Y$  are trivial). The map  $p$  is étale by the choice of orbifold indices. An étale covering is a fibration in the étale topology, and the fiber over a point (say a general point) is a one-point DM-stack, hence of the form  $BG$  for a finite group  $G$ . We obtain a long exact sequence in homotopy. If  $X$  is elliptic or hyperbolic, then by definition the same is true of  $Y$ . Thus, the universal covering of  $Y$  is  $\mathbb{C}$  or a disk [BN06], so  $\pi_2(Y) = 0$ . In particular, the long exact homotopy sequence of homotopy groups gives an extension as in the statement of the lemma. The fact that this extension uniquely determines the fibration  $p : X \rightarrow Y$ , plus the statement about maps to  $X$ , will be left as exercises in non-abelian cohomology [Bre94].  $\square$

We will see below that the notion of DM-curve gives a handy way of dealing with the difference between  $\text{SL}(2)$  and  $\text{PSL}(2)$  for factorization statements. In particular, since we are mainly using the notion of DM-curve for that purpose, it mostly suffices to consider the case where the group  $G$  in Lemma 2.5 is of order two.

Putting together the above information we obtain the following.

LEMMA 2.6. *If  $X$  is an elliptic or hyperbolic DM-curve, then there is an étale Galois covering  $Z \rightarrow X$  with  $Z$  a smooth curve; in particular,  $X$  is a quotient of  $Z$  by the Galois group  $G$ —note that the action of  $G$  on  $Z$  might not be faithful; this is how we obtain non-trivial generic stabilizers. As described above we have an extension*

$$1 \rightarrow \pi_1(Z) \rightarrow \pi_1(X) \rightarrow G \rightarrow 1.$$

*Proof.* Use Lemma 2.5 to express  $X$  as a gerb over an orbicurve  $Y$ . From the long exact sequence we see that  $Y$  has infinite fundamental group, so we are not in the degenerate cases of Lemma 2.3. Thus, there is an étale covering map  $V \rightarrow Y$  from a smooth curve to  $Y$ . Pulling back, we obtain a gerb  $U := X \times_Y V$  over  $V$ . The gerb corresponds to an extension

$$1 \rightarrow H \rightarrow \pi_1(U) \rightarrow \pi_1(V) \rightarrow 1.$$

Let  $U' \rightarrow U$  denote the covering corresponding to the map  $\pi_1(U) \rightarrow \text{Aut}(H)$ . This is again a gerb corresponding to a central extension

$$1 \rightarrow Z(H) \rightarrow \pi_1(U') \rightarrow \pi_1(V') \rightarrow 1.$$

The central extension is classified by an element of  $H^2(V', Z(H))$ . There is a finite étale covering space  $V'' \rightarrow V'$  such that the pullback of the element is zero, that is the extension splits. This gives an étale covering morphism from  $V''$  back down to  $X$ . Galois completion yields the curve  $Z$ .  $\square$

Owing to this lemma, and the fact that the other stacks we consider (Shimura modular stacks) are, by construction, analytic quotients and are seen to be algebraic quotients by looking at moduli spaces for objects with level structure, it is possible to restrict our attention to DM-stacks which are quotients.

### 2.3 Finiteness

We prove a finiteness statement which is the key to our method of proof of the implication ‘non-factoring implies rigid’ in §6. It has also independently appeared in Delzant’s preprint [Del07] with a more conceptual proof.

LEMMA 2.7. *Given  $g$  and  $b$ , there is an integer  $N$  giving the following bound. For any hyperbolic orbicurve  $X$ , and any regular curve  $Y$  with a non-constant map  $f : Y \rightarrow X$ , assume that the projective compactification  $\bar{Y}$  of  $Y$  has genus at most  $g$ , and the number of points in  $\bar{Y} - Y$  is at most  $b$ . Then for any point  $P$  in the image of  $f$ , the orbifold index of  $X$  at  $P$  is at most  $N$ .*

*Proof.* We may assume that  $X$  is compact. Indeed, if not then we could add in orbifold structures at the points at infinity of  $\bar{X} - X^{\text{coarse}}$ . Note that we are not asking in the hypothesis that the map  $f$  be surjective, only that the point  $P$  in question be in the image. If we fix an orbifold index of at least seven at these additional points, then the new orbifold will also be hyperbolic (see [BN06]), so we may assume that  $X$  is compact.

Fix  $f : Y \rightarrow X$ , and points  $Q \in Y$  and  $P = f(Q) \in X$ . Let  $n$  be the orbifold index of  $X$  at  $P$ . Denote by  $\bar{f}$  the extension of  $f$  to  $\bar{Y} \rightarrow \bar{X}$  (where  $\bar{X}$  denotes the usual curve which is the projective completion of  $X^{\text{coarse}}$ ).

The ramification degree of  $f$  at  $Q$  is divisible by  $n$ , so the contribution to the Hurwitz formula from  $Q$  is at least  $n - 1$ . If  $g(\bar{X}) \geq 1$ , we have

$$2g(\bar{Y}) - 2 \geq d(2g(\bar{X}) - 2) + (n - 1) \geq n - 1$$

which gives the bound in question.

The remaining problem is to treat the case  $\bar{X} = \mathbb{P}^1$ . Since  $X$  is hyperbolic, there are at least two other orbifold points  $P_1$  and  $P_2$  in addition to  $P_0 := P$ . Let  $n_1$  and  $n_2$  denote their orbifold indices.



Let  $d$  denote the degree of the map  $\bar{f}$ . If  $Q'$  is a point in  $\bar{f}^{-1}(P_i)$ , then either  $Q' \in Y$ , in which case the ramification degree  $r_{Q'}$  is divisible by  $n_i$  (if  $i = 0$  we are denoting  $n_0 := n$ ); or otherwise  $Q' \in \bar{Y} - Y$ . There are at most  $d/n_i$  points in the first case, and at most  $b$  points in the last case.

The Hurwitz contribution for the fiber over  $P_i$  is bounded:

$$\sum_{\bar{f}(Q')=P_i} (r_{Q'} - 1) = d - \#(\bar{f}^{-1}(P_i)) \geq d - b - d/n_i.$$

Therefore, the Hurwitz formula gives

$$2g(\bar{Y}) - 2 \geq -2d + 3d - 3b - d\left(\frac{1}{n_0} + \frac{1}{n_1} + \frac{1}{n_2}\right).$$

For  $n_0 = n$ , and noting that at least one of  $n_1$  or  $n_2$  are at least three so  $1/n_1 + 1/n_2 \leq 5/6$ , this gives a bound

$$\frac{1}{n} \geq \frac{2 - 2g(\bar{Y}) - 3b}{d} + \left(1 - \frac{1}{n_1} - \frac{1}{n_2}\right) \geq \frac{2 - 2g(\bar{Y}) - 3b}{d} + \frac{1}{6}.$$

If  $n \leq 6$ , then we have our bound so we can assume that  $n \geq 7$ . Thus,

$$\frac{3b + 2g(\bar{Y}) - 2}{d} \geq \frac{1}{42}, \quad \text{that is, } d \leq 42(3b + 2g(\bar{Y}) - 2).$$

This gives a bound for  $d$ , hence, for  $n \leq d$ .

Assume now that  $n_1 = n_2 = 2$ . In this case there must be a fourth orbifold point, because the orbifolds with indices  $(2, 2, a)$  are spherical for any  $a$ . Adding in the fourth point to the above calculations the bound becomes

$$\frac{1}{n} \geq \frac{2 - 2g(\bar{Y}) - 4b}{d} + \left(2 - \frac{1}{n_1} - \frac{1}{n_2} - \frac{1}{n_3}\right) \geq \frac{2 - 2g(\bar{Y}) - 4b}{d} + \frac{1}{2}.$$

We can assume that  $n \geq 3$ , thus

$$\frac{4b + 2g(\bar{Y}) - 2}{d} \geq \frac{1}{6}, \quad \text{that is, } d \leq 6(4b + 2g(\bar{Y}) - 2).$$

Again this gives a bound for  $n \leq d$ . □

We now obtain our finiteness result, cf. [Del07, Theorem 2].

**PROPOSITION 2.8.** *Suppose that  $X$  is a connected smooth quasiprojective variety. The set of isomorphism classes of pairs  $(Y, f)$ , where  $Y$  is an hyperbolic orbicurve and  $f : X \rightarrow Y$  is a surjective algebraic morphism, is finite.*

*Proof.* Let  $\bar{X}$  be a compactification such that the complementary divisor has normal crossings. We can find a projective embedding of  $\bar{X}$  and choose a connected family of complete intersection curves in  $\bar{X}$  passing through any point. We may assume that through any point there is a smooth curve in the family. Let  $g$  be the genus of the curves in the family, and let  $b$  be the maximal number of intersection points of a curve with the divisor  $D := \bar{X} - X$ . We may also assume that for any point  $Q \in X$  there is a curve  $\bar{C}$  in the family passing through  $Q$  and intersecting the divisor  $D$  transversally. Let  $C := \bar{C} \cap X$ , and note that  $\pi_1(C) \rightarrow \pi_1(X)$  is surjective.

Suppose that  $f : X \rightarrow Y$  is a surjective morphism to a hyperbolic orbicurve, and suppose that  $P$  is a point in the base. Choose a point  $Q \in X$  with  $f(Q) = P$  and choose a curve  $C$  in the above family passing through  $Q$ . The map  $f|_C$  is non-constant since the map on fundamental groups is non-trivial because  $\pi_1(C) \rightarrow \pi_1(X)$  is surjective and  $\pi_1(X) \rightarrow \pi_1(Y)$  has image of finite index in an infinite group. The point  $P$  is in the image of  $f|_C$ , so Lemma 2.7 provides a bound for the orbifold

index of  $Y$  at  $P$  in terms of the bounds  $g$  and  $b$  for the family of curves  $C$ . Thus, we obtain a bound  $N$  depending only on  $X$ , for the orbifold indices of points in  $Y$ . Assume that  $N \geq 7$ .

Note also that the genus of  $\bar{Y}$  is bounded by the genus  $g$  of the curves  $\bar{C}$  in the family. The number of points at infinity (that is, the cardinality of  $\bar{Y} - Y^{\text{coarse}}$ ) is bounded by the bound  $b$  for the curves  $C$ . Finally, an argument similar to that of Lemma 2.7 provides a bound for the number of orbifold points in  $Y$ . Thus, there are only finitely many possible combinatorial types for the target orbicurve  $Y$ .

For the remainder of the argument, we can fix a combinatorial type for a hyperbolic orbicurve  $Y$ . There is an étale Galois covering  $Z \rightarrow Y$  such that  $Z$  is a curve (without orbifold structure). We may also assume that the projective completion  $\bar{Z}$  has genus  $g_Z \geq 2$ . Let  $G$  denote the Galois group of  $Z/Y$ . The data of  $g_Z$  and the group may be determined by the combinatorial type of  $Y$  up to finitely many choices, so we can view them as being fixed.

Now if  $f : X \rightarrow Y$  is any map, the pullback  $X \times_Y Z$  is a Galois covering of  $X$  with group  $G$ . There are only finitely many possibilities for this covering, so we may view it as being a fixed covering  $V \rightarrow X$  with Galois action of  $G$ . The map  $f$  corresponds to a  $G$ -equivariant map from  $V$  to  $Z \subset \bar{Z}$  where  $\bar{Z}$  is a variable Riemann surface of genus  $g_Z$ . Once the map  $V \rightarrow \bar{Z}$  is fixed, the image open set  $Z$  and the action of  $G$  on  $Z$  are determined.

We are now reduced to a classical statement (see [KO75] and [Lan62, p. 137]): given a quasiprojective variety  $V$  there are only finitely many isomorphism classes of submersive maps to a compact Riemann surface of genus  $g_Z \geq 2$ . □

### 3. Factorization

Superrigidity theory says that under certain circumstances, varieties with the same topology have to be isomorphic. The notion of factorization, which extends this idea to maps that are not isomorphisms, is by now classical too [Bea92, Cam94, Cat96, Cor91, Del06, Eys04, Gro89, GS92, Kat94, Kat97, KR98, Kli03, Kol93, Sim91, Siu80, Zuo94].

Suppose that  $X$  is a smooth quasiprojective variety and  $\rho : \pi_1(X, x) \rightarrow G$  is a representation into some group  $G$ . A *factorization* is a triple  $(Y, f, \psi)$  where  $Y$  is a DM-stack,  $f : X \rightarrow Y$  is an algebraic map, and  $\psi : \pi_1(Y, f(x)) \rightarrow G$  is a representation such that  $\rho = \psi \circ \pi_1(f)$ . In this case we say that  $\rho$  *factors through*  $f : X \rightarrow Y$  or just *through*  $Y$ . If  $\pi_1(f)$  is not surjective, then the representation  $\psi$  might not be unique.

We say that  $\rho$  *projectively factors through*  $f : X \rightarrow Y$  if the composed representation

$$\tilde{\rho} : \pi_1(X) \rightarrow G/Z(G)$$

to the quotient of  $G$  by its center factors through  $f$ . This terminology is intended for the case  $G = \text{SL}(2, K)$  and  $G/Z(G) = \text{PSL}(2, K)$ .

For example we say that a representation  $\rho$  *factors through a DM-curve* if there exists a map  $f : X \rightarrow Y$  to a smooth DM-curve and a factorization of  $\rho$  through  $f$ .

Recall that an orbicurve is a DM-curve such that the generic stabilizer is trivial. The following lemmas say that it is equivalent to speak of projective factorization through maps to orbicurves, or actual factorization through maps to DM-curves. In the remainder of the paper we use the language of factorization through DM-curves since this is easier to handle notationally speaking.

LEMMA 3.1. *Suppose that  $X$  is a smooth quasiprojective variety and  $\rho : \pi_1(X, x) \rightarrow \text{SL}(2, K)$  a representation with  $K$  a field. Suppose that  $\rho$  projectively factors through a map  $f : X \rightarrow Y$  to an orbicurve, that is, there is a factorization  $(Y, f, \psi)$  for the representation  $\tilde{\rho}$  into  $\text{PSL}(2, K)$ . Then there is a structure of DM-curve  $Y'$  mapping to  $Y$  and a lifting  $f' : X \rightarrow Y'$  such that  $\rho$  factors through  $f'$ .*

*Proof.* Recall from Lemma 2.4 that  $Y$  must be hyperbolic. Use  $\psi$  to pull back the central extension

$$1 \rightarrow \{\pm 1\} \rightarrow \mathrm{SL}(2, K) \rightarrow \mathrm{PSL}(2, K) \rightarrow 1.$$

We obtain a central extension of  $\pi_1(Y)$  by a group of order two. By Lemma 2.5 this corresponds to a structure of DM-curve  $Y'$  over  $Y$  and the representation  $\psi$  lifts to a representation  $\psi' : \pi_1(Y') \rightarrow \mathrm{SL}(2, K)$ . Note that the square

$$\begin{array}{ccc} \pi_1(Y') & \longrightarrow & \pi_1(Y) \\ \downarrow & & \downarrow \\ \mathrm{SL}(2, K) & \longrightarrow & \mathrm{PSL}(2, K) \end{array}$$

is cartesian. The projective factorization gives a map into the fiber product of the lower right angle of this diagram, thus it gives a map  $\pi_1(X) \rightarrow \pi_1(Y')$ . Thus, by Lemma 2.5 the map  $f : X \rightarrow Y$  lifts to a map  $f' : X \rightarrow Y'$  giving a factorization of the representation  $\rho$ .  $\square$

In the above lemma, the structure of DM curve is not unique.

LEMMA 3.2. *Suppose that  $X$  is a smooth quasiprojective variety and  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, K)$  is a Zariski-dense representation with  $K$  a field. If  $\rho$  factors through a map to a DM-curve  $f : X \rightarrow Y'$ , then letting  $Y$  be the underlying orbicurve as defined above,  $\rho$  projectively factors through  $f : X \rightarrow Y$ .*

*Proof.* Suppose that  $(Y', f', \psi')$  is a factorization for  $\rho$ . Note that  $Y'$  is hyperbolic so it is classified by the extension

$$1 \rightarrow G \rightarrow \pi_1(Y') \rightarrow \pi_1(Y) \rightarrow 1.$$

The group  $G$  is finite. The subgroup of  $\mathrm{SL}(2, K)$  which normalizes the image  $\psi'(G)$  is an algebraic subgroup containing the image of  $\rho$ . By the Zariski-density hypothesis, this group is the whole of  $\mathrm{SL}(2, K)$ , that is,  $\psi'(G)$  is a finite normal subgroup, hence contained in the center. Thus, the projection  $\tilde{\rho}$  into  $\mathrm{PSL}(2, K)$  is trivial on  $G$  so we obtain a factorizing representation  $\psi : \pi_1(Y) \rightarrow \mathrm{PSL}(2, K)$ .  $\square$

COROLLARY 3.3. *Suppose that  $X$  is a smooth quasiprojective variety and  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, K)$  a Zariski-dense representation with  $K$  a field. Then  $\rho$  factors through a map to a DM-curve, if and only if it projectively factors through a map to an orbicurve.*

LEMMA 3.4. *If  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, K)$  is a Zariski-dense representation which factors through a map  $f : X \rightarrow Y$  to a DM-curve, then  $Y$  is a hyperbolic DM-curve.*

*Proof.* This follows from Lemma 2.4.  $\square$

### 3.1 Invariance under open sets and finite coverings

LEMMA 3.5. *Suppose that  $U \subset X$  is a Zariski-open set in a smooth quasiprojective variety. Suppose that  $K$  is an infinite field and  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, K)$  is a Zariski-dense representation such that the restriction  $\rho|_U$  factors through a DM-curve. Then  $\rho$  itself factors through a DM-curve.*

*Proof.* It is easier to use Corollary 3.3 to replace factorization through a DM-curve, by projective factorization through an orbicurve in the statement. Thus, we look at the representation  $\tilde{\rho}$  projected into  $\mathrm{PSL}(2, K)$ .

Let  $f : U \rightarrow Y$  be the factorization map to an orbicurve  $Y$ . We can restrict the open set  $U$ , so we can suppose that  $Y$  is a smooth curve. Let  $\bar{Y}$  denote the smooth projective compactification of  $Y$ . For  $P \in \bar{Y} - Y$  consider the monodromy transformation for  $\tilde{\rho}_Y$  around  $P$ . If it is of infinite order, then leave out the point  $P$ . If it is of some finite order  $n$ , then add an orbifold point of order

$n$  at  $P$ . This gives us an orbicurve partial completion  $Y'$  of  $Y$  with a representation  $\tilde{\rho}_{Y'}$  restricting to  $\tilde{\rho}_Y$  on  $Y$ .

Now we claim that the map  $f$  extends to a map  $f' : X \rightarrow Y'$ . For this, note that there is a birational modification  $p : X'' \rightarrow X$  containing  $U \subset X'$  as the complement of a divisor with normal crossings, and a map  $f'' : X'' \rightarrow \bar{Y}$ . On the other hand, for any component of the complementary divisor in  $X''$ , the monodromy around that component is trivial for the representation  $\tilde{\rho}$ . This implies that any divisor component maps to a point of  $Y'$  (that is, maps to a point in  $\bar{Y}$  where the monodromy is not of infinite order). Thus, we obtain a map  $X'' \rightarrow (Y')^{\text{coarse}}$ .

By looking at the order of the monodromy around any complementary divisor component, we obtain a lifting to a map  $X'' \rightarrow Y'$ . Suppose that  $P \in (Y')^{\text{coarse}}$  is an orbifold point with index  $n$ . The inverse image  $(f'')^{-1}(P) = \sum_i r_i B_i$  is a divisor in  $X''$ . If  $B_i$  is a component with multiplicity  $r_i$ , then  $r_i$  is divisible by  $n$ , because  $\tilde{\rho}_Y$  has monodromy of order  $n$  around  $P$  and the pullback  $f^*(\tilde{\rho}_Y)$  extends to  $X$  so it extends across the component  $B_i$ . Now each  $r_i$  is divisible by  $n$  so the divisor  $(f'')^{-1}(P)$  is divisible by  $n$ , which says that  $X'' \rightarrow (Y')^{\text{coarse}}$  lifts to a map into  $Y'$  locally at  $P$ .

Finally, we can go to a model  $X^{(k)}$  birational over  $X''$  and with the property that  $X^{(k)}$  is obtained from  $X = X^{(0)}$  by a sequence of blow-ups  $\varphi_{i+1} : X^{(i+1)} \rightarrow X^{(i)}$  with smooth centers  $V^{(i+1)}$ . Working by downward induction on  $i$ , we descend the map  $X^{(k)} \rightarrow Y'$  to a map  $X \rightarrow Y'$  as follows. Suppose that we are given  $f_{i+1} : X^{(i+1)} \rightarrow Y'$ . For any point on the center of the blow-up  $P \in V^{(i+1)} \subset X^{(i)}$  the inverse image  $\varphi_{i+1}^{-1}(P)$  is a connected variety such that

$$f_{i+1}^*(\tilde{\rho}_{Y'})|_{\varphi_{i+1}^{-1}(P)} = \rho|_{\varphi_{i+1}^{-1}(P)}$$

is trivial. Therefore,  $f_{i+1}$  maps  $\varphi_{i+1}^{-1}(P)$  to a single point in  $Y'$ . Thus,  $f_{i+1}$  descends to a map

$$X^{(i)} \rightarrow (Y')^{\text{coarse}},$$

and looking again at the monodromy around orbifold points as in the previous paragraph, we find that it lifts to

$$X^{(i)} \rightarrow Y'.$$

Continuing by downward induction on  $i$  we eventually come to a map  $f' : X \rightarrow Y'$  as required. This constitutes a factorization of the representation  $\rho$ , indeed  $(f')^*(\tilde{\rho}_{Y'}) = \rho$  because this is true over the Zariski-open dense set  $U$ . □

**LEMMA 3.6.** *Suppose that  $p : Z \rightarrow X$  is a quasifinite map of smooth quasiprojective varieties. Suppose that  $K$  is an infinite field and  $\rho : \pi_1(X, x) \rightarrow \text{SL}(2, K)$  is a Zariski-dense representation such that the pullback  $\rho|_Z$  factors through a DM-curve. Then  $\rho$  itself factors through a DM-curve.*

*Proof.* This is an easy special case of the Shafarevich factorization, see [Kol93] as well as [Cam94, Eys04, Kat97]. We handle it explicitly here.

For constructing the factorization, we may pass to a Zariski-open set of  $X$  and then invoke Lemma 3.5. In particular, we may assume that  $p$  is a finite map.

Let  $f : Z \rightarrow Y$  be the factorization map. By Lemma 3.4,  $Y$  is an hyperbolic DM-curve. Therefore, there exists a finite covering  $Y' \rightarrow Y$  by a smooth curve. Replacing  $Z$  by its fiber product with  $Y'$ , we find that we may, in fact, assume that  $\rho|_Z$  factors through a smooth curve. Thus, take now this situation and assume that  $f : Z \rightarrow Y$  is the factorization map to a smooth curve. We may also take the Stein factorization so we can assume that the general fiber of  $f$  is connected. Remove from  $X$  the closure of the image of all of the non-smooth or reducible fibers. Thus, we may assume that all of the fibers of  $f$  are irreducible and smooth.

We can assume that  $Z$  is Galois over  $X$  with group  $G$ . We claim that  $G$  permutes the fibers of  $f$ . This is because a fiber is characterized as a maximal connected smooth curve such that the

restriction of  $\rho$  is trivial. Therefore, the action of  $G$  extends to an action on  $Y$  compatible with the map  $f$ . Let  $V := Y/G$  be the DM-curve quotient. We obtain a factorization map  $X \rightarrow V$  for  $\rho$ .

The fact that the action of  $G$  on  $Z$  preserves  $\rho$  (because  $\rho$  is a pullback from  $X$ ) allows us to define an action on  $\rho_Y$  covering the action on  $Y$ , so  $\rho_Y$  descends to a representation  $\rho_V$  on  $V$  whose pullback to  $X$  is  $\rho$ .  $\square$

### 3.2 Invariance under field extension

Suppose that  $f : X \rightarrow Y$  is a map. Let  $K$  denote the kernel of the induced map  $\pi_1(X, x) \rightarrow \pi_1(Y, f(x))$ . A representation  $\rho : \pi_1(X, x) \rightarrow G$  factors through  $f$  if and only if the restriction of  $\rho$  to  $K$  is trivial. This immediately gives the following statement.

LEMMA 3.7. *Suppose that  $\alpha : G \hookrightarrow H$  is an injection of groups. A representation  $\rho : \pi_1(X, x) \rightarrow G$  factors through a map  $f : X \rightarrow Y$ , if and only if the composed representation  $\alpha \circ \rho : \pi_1(X, x) \rightarrow H$  factors through  $f$ .*

We have the following independence of factorization with respect to change of coefficient field, as a corollary.

COROLLARY 3.8. *Suppose that  $K \subset L$  is a field extension. Then a representation  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(r, K)$  factors (respectively projectively factors) through a map  $f : X \rightarrow Y$  if and only if the representation obtained by extension of coefficients  $\rho_L : \pi_1(X, x) \rightarrow \mathrm{SL}(r, L)$  factors (respectively projectively factors) through  $f$ .*

## 4. Construction of a pluriharmonic mapping to the Bruhat–Tits tree

An element  $A \in \mathrm{PSL}(2, K)$  is said to be *unipotent* if its representatives in  $\mathrm{SL}(2, K)$  are conjugate to upper-triangular matrices with  $\pm 1$  on the diagonal. An element is *quasiunipotent* if there is some  $n$  such that  $A^n$  is unipotent. We consider the following situation, stated as a hypothesis for future reference.

HYPOTHESIS 4.1. *Fix a quasiprojective variety  $X$  with compactification  $\overline{X}$  whose complementary divisor  $D$  is a union of smooth irreducible components  $D_i$  intersecting at normal crossings. Let  $x$  denote a basepoint in  $X$ .*

*Let  $\mathcal{O}_K$  be a complete local ring with fraction field  $K$  and residue field  $F$ ; we assume that  $F$  is finite. Suppose that we have a representation  $\rho : \pi_1(X, x) \rightarrow \mathrm{PSL}(2, K)$  such that the monodromy transformations around the  $D_i$  are quasiunipotent. We also assume that the monodromy transformations around the  $D_i$  are not the identity (that is, that  $X$  is a maximal open subset of  $\overline{X}$  of definition for the representation  $\rho$ ). We assume that the Zariski closure of the image of  $\rho$  is all of  $\mathrm{PSL}(2, K)$ , and we assume that the image is not contained in any compact subgroup.*

In the next section, we prove Theorem 5.13 saying that under the above hypothesis the representation  $\rho$  factors through a DM-curve. In general Hypothesis 4.1 will be in effect throughout this section and the next.

The basic technique is to let  $\mathrm{PSL}(2, K)$  act on the *Bruhat–Tits tree*  $\mathcal{T}$ , and to choose an equivariant pluriharmonic map  $\tilde{X} \rightarrow \mathcal{T}$  via the theory of Gromov and Schoen [GS92]. The first step, not totally trivial in the quasiprojective case, is to choose an initial map having finite energy on the fundamental region  $\mathrm{Reg}(\tilde{X}/X)$ . By [GS92] this deforms to a pluriharmonic map which is studied in § 5.

This theory has already been treated by Jost and Zuo in [JZ00]. They treated the case of harmonic maps to buildings of arbitrary rank. In the case of trees which was treated specifically

in [GS92], and which is by now pretty well known, many parts of the argument are easier to understand so it seems worthwhile to recall them here. Also, we feel that the discussion of the choice of initial map on [JZ00, p. 12] needs some amplification.

Looking at harmonic mappings to trees or more generally buildings, mirrors much of the study of harmonic bundles which has been extended to the quasiprojective case [Biq96, Car92, JZ96, Li00, LN99, LN01, Moc02, Moc07, NR01b, Sim90]. Other related techniques are also available [Del07, Gro89, NR01a, NR08, NR06].

### 4.1 The Bruhat–Tits tree

Recall the construction of the *Bruhat–Tits tree* associated to  $K$ , extensively studied by Bass [Bas80] and Serre [Ser80, Ser77]. There is a uniformizing parameter  $t \in \mathcal{O}_K$  which generates the maximal ideal. The residue field  $F := \mathcal{O}_K/(t)$  is assumed to be finite, with  $q = p^k$  elements. A *lattice* is a finitely generated  $\mathcal{O}_K$ -submodule  $M \subset K^2$  which spans  $K^2$  over  $K$ . It follows that  $M \cong \mathcal{O}_K^2$ .

Say that two lattices  $M_1$  and  $M_2$  are *equivalent* if one is a scalar multiple of the other by an element of  $K^*$ . The vertices of the *Bruhat–Tits tree* are defined to be the equivalence classes of lattices. Two lattices are *adjacent* if after choosing appropriate representatives up to equivalence, one is contained in the other and if the quotient is an  $\mathcal{O}_K$ -module of length one isomorphic to  $F$ . The edges of the Bruhat–Tits tree are defined as connecting lattices which are adjacent in the above sense. See [Ser77]. There are  $q + 1$  edges at each vertex.

Metetrize the tree by assigning a linear metric of length one to each edge.

The group  $SL(2, K)$  acts isometrically on  $\mathcal{T}$  via its action on  $K^2$ , thus  $g \in SL(2, K)$  sends a lattice  $M$  to  $gM$ . The center  $\{\pm 1\}$  acts trivially on  $\mathcal{T}$ , so the action factors through the group  $PSL(2, K)$ .

Choosing a basis for a lattice  $M$  gives an isomorphism between  $PSL(2, \mathcal{O}_K)$  and the subgroup of elements which fix  $M$ , in particular the latter group is compact.

Conversely any compact subgroup  $H \subset PSL(2, K)$  fixes a lattice. Indeed, the action on the tree is continuous, so the orbit of a point under the action of  $H$  is compact. The convex subset spanned by this orbit is  $H$ -invariant and compact, so the set of endpoints is finite and the barycenter of this finite set is a fixed point for  $H$ .

Recall that an *end* of  $\mathcal{T}$  is an equivalence class of injective isometric paths  $\mathbb{R}_{\geq 0} \rightarrow \mathcal{T}$ , under the equivalence relation that two paths are said to be equivalent if they remain a bounded distance apart or, equivalently, if their image subsets coincide after a finite time. Recall also that a parabolic subgroup of  $PSL(2, K)$  is a subgroup conjugate to the group of upper-triangular matrices. Such a subgroup is uniquely determined by its corresponding rank-one subspace  $L \subset K^2$ .

LEMMA 4.2. *The subgroup of elements fixing an end is a parabolic subgroup, and a parabolic subgroup fixes a unique end. In this way the set of ends of  $\mathcal{T}$  is naturally in one-to-one correspondence with the set of parabolic subgroups of  $PSL(2, K)$ .*

See [Bas80, Ser80, Ser77] for the proof.

### 4.2 Quasiunipotent matrices

Let  $\overline{K}$  denote the algebraic closure of  $K$ . A matrix  $A \in PSL(2, K)$  is quasiunipotent if and only if the image in  $PSL(2, \overline{K})$  is conjugate to an upper-triangular matrix whose diagonal entries are roots of unity. The diagonal entries are  $\alpha, \alpha^{-1}$  with  $\alpha^n = 1$  for some  $n$ .

Let  $Z(A) \subset PSL(2, K)$  be the centralizer of  $A$ , that is the group of matrices in  $PSL(2, K)$  which commute with  $A$ . It is an algebraic group and its extension of scalars to  $\overline{K}$  is the same as the centralizer of  $A$  in  $PSL(2, \overline{K})$ . The centralizer will play an important role in our construction below. We divide into three cases according to the type of  $Z(A)$  up to conjugacy.

LEMMA 4.3. *Suppose that  $A \in \text{PSL}(2, K)$  is a non-trivial quasiunipotent matrix. Then one of three possibilities holds.*

- (i) Unipotent: *the matrix  $A$  is conjugate to an upper-triangular matrix with  $\pm 1$  on the diagonal, and the centralizer  $Z(A)$  is conjugate to the unipotent group of all such matrices.*
- (ii) Split torus: *the matrix  $A$  is conjugate to a diagonal matrix with entries  $\alpha^{\pm 1}$  where  $\alpha \neq \pm 1$  is a non-trivial root of unity, and the centralizer  $Z(A)$  is conjugate to the split torus  $\mathbb{G}_m$  of diagonal matrices.*
- (ii) Non-split torus: *the matrix  $A$  has eigenvalues  $\alpha^{\pm 1}$  which are not in  $K$ , and the centralizer  $Z(A)$  is a non-split torus which becomes conjugate in  $\text{PSL}(2, \overline{K})$  to the group of diagonal matrices.*

*Proof.* Over  $\overline{K}$  the matrix can be put in Jordan normal form. If it has distinct eigenvalues, then they are either in  $K$ , which gives the split torus case, or not in  $K$ , giving the non-split torus case. If both eigenvalues are the same, that is to say  $\pm 1$ , and if the matrix is non-trivial in  $\text{PSL}(2, \overline{K})$  then it is conjugate to  $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ . In this case, there is a unique eigenvector, and the dimension of  $\ker(A - 1)$  is invariant under field extension. Thus,  $A$  has an eigenvector over  $K$  so it can be conjugated to a unipotent upper-triangular matrix in  $\text{PSL}(2, K)$ . This gives the unipotent case.  $\square$

LEMMA 4.4. *Suppose that  $A$  and  $B$  are commuting quasiunipotent matrices, both non-trivial in  $\text{PSL}(2, K)$ . Then  $Z(A) = Z(B)$ .*

*Proof.* It suffices to do this in  $\text{PSL}(2, \overline{K})$ . Thus, we may assume that  $A$  is either diagonal with distinct eigenvalues, or  $\pm 1$  times an upper-triangular matrix. Respectively, we see that  $B$  has to be diagonal or  $\pm 1$  times an upper-triangular matrix. Again, in these cases the centralizers in  $\text{PSL}(2, \overline{K})$  are either the group of diagonal matrices or the group of projectively upper-triangular matrices, respectively. The centralizers of  $A$  and  $B$  in  $\text{PSL}(2, \overline{K})$  are the same in either case. Formation of the centralizer commutes with field extension, so this shows that  $Z(A) = Z(B)$  in  $\text{PSL}(2, K)$ .  $\square$

### 4.3 Local study at infinity

Recall that part of our assumption is that the monodromy around all components of the divisor  $D$  is non-trivial in  $\text{PSL}(2, K)$ . Thus, the connected components of  $D$  form well-defined subsets depending on  $\rho$ . The choice of initial map is a local problem near connected components of  $D$ . In view of this, for simplicity we assume that  $D$  is connected for most of the remainder of this section.

Let  $D_j$  denote the irreducible components of  $D$ , which we are assuming are smooth. Let  $N_j$  denote a tubular neighborhood of  $D_j$  in  $\overline{X}$  and let  $N_j^*$  denote the complement of  $D$ , that is  $N_j^* = N_j \cap X$ . Let  $D_{ij} := D_i \cap D_j$ , etc., and let  $N_{ij} := N_i \cap N_j$  which we assume is a tubular neighborhood of  $D_{ij}$  (with polydisk transversal section). Let  $N := \bigcup_j N_j$  be the tubular neighborhood of  $D$  and  $N^* := N \cap X$ .

For now, assume that we have chosen a basepoint  $x \in N^*$ . Let  $\tilde{N}^*$  denote the resulting universal covering of  $N^*$ . Let  $\text{Reg}(\tilde{N}^*/N^*) \subset \tilde{N}^*$  denote a closed fundamental region for  $N^*$ , as will be chosen more specifically further along (Lemma 4.8). Let  $\pi_1(N^*)$  denote the fundamental group of  $N^*$  based at  $x$ , also viewed as the group of deck transformations of  $\tilde{N}^*$ .

Let  $\gamma_j$  be a path going around the divisor component  $D_j$ . Recall that we are assuming that  $\rho(\gamma_j)$  is quasiunipotent and non-trivial in  $\text{PSL}(2, K)$ . We assume that the basepoint  $x$  is near  $D$  and that the paths used to define the  $\gamma_j$  stay within the tubular neighborhood  $N$  of  $D$ .

LEMMA 4.5. *Assuming that Hypothesis 4.1 as well as connectedness of  $D$  hold and with the above notation, let  $Z(\rho(\gamma_j))$  be the centralizer in  $\text{PSL}(2, K)$  of the matrix  $\rho(\gamma_j)$ . Then the  $Z(\rho(\gamma_j))$  are all the same group for different  $j$ .*

*Proof.* We may assume that the basepoint  $x$  is near one of the components  $D_0$ . Write

$$\gamma_j = \eta \xi_j \eta^{-1}$$

where  $\xi_j$  is a path going around  $D_j$  from a nearby basepoint  $x_j$ , and  $\eta$  is a path going from  $x$  to  $x_j$ . Decompose  $\eta$  into paths

$$\eta = \eta_1 \cdots \eta_k$$

where for any  $a = 1, \dots, k$ , the piece  $\eta_a$  is a path staying in  $N_{i_a}$ . Let  $\zeta_a$  be a loop going from the endpoint of  $\eta_a$  around the divisor component  $D_{i_a}$  and let  $\zeta'_a$  be a loop going around the divisor component  $D_{i_{a+1}}$ . Note that the endpoint of  $\eta_a$  is in  $N_{i_{a+1}}$  so we can look at both of these loops.

Commutativity of the fundamental group of the product of punctured polydisks says that  $\zeta_a$  and  $\zeta'_a$  commute. This step corresponds to the phrase on lines 4–7 of page 12 of [JZ00] and in a certain sense the rest of our discussion above and below consists of filling in the surrounding details.

For any  $i$  the loop around the divisor  $D_i$  is in the center of  $\pi_1(N_i^*)$ , as can be seen by restricting to a Zariski-open set of  $D_i$  over which the normal bundle is trivialized, and noting that the fundamental group of the open set surjects to the original one. This implies that

$$\zeta'_a = \eta_{a+1} \zeta_{a+1} \eta_{a+1}^{-1}.$$

Set

$$\tau_a := \eta_1 \cdots \eta_a \zeta_a \eta_a^{-1} \cdots \eta_1^{-1}.$$

Combining with commutativity of  $\zeta_a$  and  $\zeta'_a$  from the previous paragraph we conclude that  $\tau_a$  and  $\tau_{a+1}$  commute. By our hypothesis on  $D$ , all of the  $\rho(\tau_a)$  are non-trivial quasiunipotent matrices in  $\mathrm{PSL}(2, K)$ . By Lemma 4.4 we find

$$Z(\rho(\tau_a)) = Z(\rho(\tau_{a+1})).$$

It follows by induction on  $a$  that the centralizers  $Z(\rho(\tau_a))$  are the same for all  $a$ . We can extend some of our notation to the boundary cases  $a = 0$  and  $a = k$  as is left to the reader. With this,  $\tau_0 = \gamma_0$  whereas  $\tau_k = \gamma_j$ , and we obtain

$$Z(\rho(\gamma_0)) = Z(\rho(\gamma_j)).$$

This proves the lemma. □

In view of this lemma, let  $Z \subset \mathrm{PSL}(2, K)$  denote the centralizer. We can assume by an appropriate change of basis that it is either the group of unipotent upper-triangular matrices, the group of diagonal matrices, or a non-split torus, see Lemma 4.3. Let  $NZ \subset \mathrm{PSL}(2, K)$  denote its normalizer. In the unipotent case,  $NZ$  is the parabolic subgroup  $P$  of upper-triangular matrices; in the split torus case,  $NZ/Z$  is the Weyl group of order two transposing the two standard basis elements of  $K^2$ , and in the non-split torus case,  $NZ/Z$  has order one or two by mapping it into the Weyl group over the algebraic closure.

LEMMA 4.6. *Let  $NZ$  denote the normalizer of the subgroup  $Z$  which is the centralizer of any  $\rho(\gamma_j)$ . Then the image by  $\rho$  of the fundamental group  $\pi_1(N^*)$  lies inside  $NZ$ . Furthermore, the representation gives a map of exact sequences.*

$$\begin{array}{ccccccc} 1 & \longrightarrow & \pi_1(N^* \times_D \tilde{D}) & \longrightarrow & \pi_1(N^*) & \longrightarrow & \pi_1(D) \longrightarrow 1 \\ & & \downarrow & & \downarrow & & \downarrow \\ 1 & \longrightarrow & Z & \longrightarrow & NZ & \longrightarrow & NZ/Z \longrightarrow 1 \end{array}$$



*Proof.* Let  $D_0$  be the irreducible component of  $D$  near the basepoint  $x$ , and let  $\gamma_0$  be the loop going from  $x$  around  $D_0$ . We claim that for any  $\eta \in \pi_1(N^*, x)$  then

$$Z(\rho(\eta\gamma_0\eta^{-1})) = Z(\rho(\gamma_0)).$$

The proof of this claim is exactly the same as the proof of Lemma 4.5: decompose

$$\eta = \eta_1 \cdots \eta_k,$$

fix the loops  $\zeta_a$  as previously and set

$$\tau_a := \eta_1 \cdots \eta_a \zeta_a \eta_a^{-1} \cdots \eta_1^{-1}.$$

This time we choose  $\zeta_k := \gamma_0$  which gives for endpoints  $\tau_0 = \gamma_0$  and  $\tau_k = \eta\gamma_0\eta^{-1}$ . As before  $\tau_a$  and  $\tau_{a+1}$  commute, so  $Z(\rho(\tau_a)) = Z(\rho(\tau_{a+1}))$  by Lemma 4.4. Thus,  $Z(\rho(\tau_k)) = Z(\rho(\tau_0))$  which gives the claim.

On the other hand, the centralizers of conjugate matrices are conjugate, that is

$$Z(\rho(\eta\gamma_0\eta^{-1})) = \rho(\eta)Z(\rho(\gamma_0))\rho(\eta)^{-1}.$$

Putting this together with the claim and the notation  $Z = Z(\rho(\gamma_j))$  for any  $j$ , we find that

$$\rho(\eta)Z\rho(\eta)^{-1} = Z.$$

Thus,  $\rho(\eta) \in NZ$  is in the normalizer of  $Z$ .

We finish by showing the map of exact sequences. The fact that the image of  $\rho$  on  $\pi_1(N^*)$  is contained in the normalizer of  $Z$  implies that the normal subgroup of  $\pi_1(N^*)$  generated by the  $\gamma_j$  maps into  $Z$ . However, this normal subgroup is the kernel of the map

$$\pi_1(N^*) \rightarrow \pi_1(N) = \pi_1(D),$$

as may be seen by noting that  $N$  is obtained homotopically from  $N^*$  by adding some 2-cells whose boundaries are the loops  $\gamma_j$ , and then adding only cells of dimensions at least three. Thus, the kernel of the upper exact sequence maps into  $Z$ , so  $\rho$  induces the map of exact sequences as claimed.  $\square$

**COROLLARY 4.7.** *The matrices  $\rho(\gamma_j)$  all commute with each other.*

*Proof.* They are all inside the group  $Z$ , which in the case of  $\text{PSL}(2, K)$  is commutative.  $\square$

*Remark 1.* Let  $\Sigma$  denote the incidence simplicial complex of the divisor  $D$ . This has appeared in the works of Stepanov [Ste06] and Thuillier [Thu07]. There is a vertex for each divisor component  $D_j$ , an edge for each irreducible component of  $D_{ij}$ , and higher simplices for the multiple intersections. We have a map  $N \rightarrow \Sigma$  inducing  $\pi_1(N^*) \rightarrow \pi_1(\Sigma)$ . In Lemma 4.6, we could have replaced  $D$  by  $\Sigma$ , that is to say the representation sends the kernel of  $\pi_1(N^*) \rightarrow \pi_1(\Sigma)$  into the center  $Z$ . We do not need this for our proof, and the end theorem of the paper shows that the situation we are considering here does not really arrive in practice; however, this remark might be interesting from a geometrical point of view in related situations such as [Del07].

#### 4.4 Local construction of an initial map: generalities

Fix a complete finite-volume Kähler metric of Poincaré type on  $X$ , see [CKS87, JY86, JZ00, SY82]. Our first problem will be to choose an initial finite-energy equivariant map to the Bruhat–Tits tree under the action of  $\text{PSL}(2, K)$ .

Composing our representation  $\rho$  with the inclusion  $\pi_1(N^*) \rightarrow \pi_1(X)$  we obtain an action denoted again by  $\rho$  of  $\pi_1(N^*)$  on the tree  $\mathcal{T}$ . We would like to construct an initial map  $\phi : \tilde{N}^* \rightarrow \mathcal{T}$ , which is  $\rho$ -equivariant and whose restriction to the fundamental region  $\text{Reg}(\tilde{N}^*/N^*)$  has finite energy.

We assume that we have chosen projections  $p_j : N_j \rightarrow D_j$  such that on  $N_{ij}$  the different projections commute and give a projection  $p_{ij}$  to  $D_{ij}$ , and similarly for higher intersections.

In order to insure the finite-energy condition, we try to construct a map which is constant in the direction of the fibers of  $N_j \rightarrow D_j$ . This would mean in particular that it is constant on the fibers of  $N_{ij} \rightarrow D_{ij}$  and similarly for multiple intersections. The discussion by Jost and Zuo in [JZ00, pp. 8–12] serves to explain why this condition yields a finite-energy map. To be more precise about this, say that two points in  $\tilde{N}^*$  (or, equally well,  $N^*$  or  $N$ ) are *projection-equivalent* if they are joined by a path which projects into  $N^*$  to a path which stays in any fiber of any projection  $p_j$  which it meets. We refer to [JZ00] for the fact that any Lipschitz and piecewise smooth map  $\phi : \tilde{N}^* \rightarrow \mathcal{T}$  which sends projection-equivalent points to the same point in  $\mathcal{T}$ , has finite energy on the fundamental region.

If we take the quotient of  $N$  by the equivalence relation of projection-equivalence we obtain a space which is homotopy-equivalent to  $D$ . Thus, by modifying the projections  $p_j$  near the multiple intersections, we can define a map  $\sigma : N \rightarrow D$  with the property that all fibers of the projections  $p_j$  go to single points in  $D_j$ . The restriction  $\sigma|_D$  will no longer be the identity, for example, points in  $N_{ij} \cap D$  have to be mapped to  $D_{ij}$ . However,  $\sigma|_D$  will be homotopic to the identity. The map  $\sigma$  induces

$$\sigma_* : \pi_1(N^*) \rightarrow \pi_1(D), \quad \tilde{\sigma} : \tilde{N}^* \rightarrow \tilde{D},$$

where  $\tilde{D} \rightarrow D$  is the universal cover corresponding to  $\sigma(x)$  as a basepoint for  $D$ .

We assume that the choice of fundamental region  $\text{Reg}(\tilde{N}^*/N^*)$  is made compatibly with a choice of  $\text{Reg}(\tilde{D}/D)$  via the map  $\sigma$ . More precisely, proceed as follows. Let  $N^+$  denote the real blow-up of the divisor  $D$  in  $N$ . It consists of adding a boundary to  $N^*$ , the boundary being homeomorphic to the boundary of a tubular neighborhood of  $D$ . The inclusion of an open dense subset  $N^* \hookrightarrow N^+$  is a homotopy equivalence. We assume that our map extends to a map

$$\sigma^+ : N^+ \rightarrow D$$

and the homotopy equivalence between  $N^*$  and  $N^+$  commutes with  $\sigma^+$ . Let  $\tilde{N}^+$  denote the universal cover of  $N^+$  corresponding to the same basepoint  $x$ , so again we have an open dense inclusion  $\tilde{N}^* \hookrightarrow \tilde{N}^+$  on which the homotopy equivalence lifts.

Choose a fundamental region

$$\text{Reg}(\tilde{D}/D) \subset \tilde{D}.$$

Now  $\sigma^+$  induces  $\tilde{\sigma}^+ : \tilde{N}^+ \rightarrow \tilde{D}$ . Consider the region

$$Q^+ := (\tilde{\sigma}^+)^{-1}(\text{Reg}(\tilde{D}/D)) \subset \tilde{N}^+,$$

and let  $Q^* = Q^+ \cap \tilde{N}^*$ .

Note that  $\pi_1(N^+) = \pi_1(N^*)$  and  $\sigma_* = (\sigma^+)_*$  maps this group to  $\pi_1(D)$ . The group  $\ker(\sigma_*) \subset \pi_1(N^*)$  acts freely on  $Q^+$ , and the quotient is

$$Q^+ / \ker(\sigma_*) = N^+ \times_D \text{Reg}(\tilde{D}/D).$$

Let  $\text{Reg}(\tilde{N}^+/N^+) \subset Q^+$  be a fundamental region for the action of  $\ker(\sigma_*)$ .

In general  $\ker(\sigma_*)$  will not be finitely generated. However,  $N^+ \times_D \text{Reg}(\tilde{D}/D)$  is homeomorphic to a finite simplicial complex, so its fundamental group is finitely generated. Thus,  $Q^+$  will usually be disconnected, a disjoint union of coverings of  $N^+ \times_D \text{Reg}(\tilde{D}/D)$  corresponding to finitely generated groups.

The fact that the quotient is a finite complex means we can choose a compact fundamental region in  $Q^+$ , indeed in one of the connected components. Thus,  $\text{Reg}(\tilde{N}^+/N^+)$  is compact and is a finite simplicial complex.

Now let  $\text{Reg}(\tilde{N}^*/N^*) := \text{Reg}(\tilde{N}^+/N^+) \cap N^*$ . Note that it is a fundamental region in  $Q^*$  for the action of  $\ker(\sigma_*)$ , surjecting to the quotient

$$Q^* / \ker(\sigma_*) = N^* \times_D \text{Reg}(\tilde{D}/D).$$

LEMMA 4.8. *The region  $\text{Reg}(\tilde{N}^*/N^*)$  constructed in this way is fundamental, that is any point of  $\tilde{N}^*$  is a  $\pi_1(N^*)$ -translate of a point in the region; also the subset  $\mathcal{A} \subset \pi_1(N^*)$  of elements  $\alpha$  such that*

$$\text{Reg}(\tilde{N}^*/N^*) \cap \alpha \cdot \text{Reg}(\tilde{N}^*/N^*) \neq \emptyset$$

is finite.

*Proof.* It suffices to do this for  $\text{Reg}(\tilde{N}^+/N^+)$  because

$$\text{Reg}(\tilde{N}^*/N^*) = \text{Reg}(\tilde{N}^+/N^+) \times_{N^+} N^*,$$

and  $\tilde{N}^* = \tilde{N}^+ \times_{N^+} N^*$  too.

The map  $\sigma_* : \pi_1(N^*) \rightarrow \pi_1(D)$  is surjective. This implies that any point in  $\tilde{N}^+$  is a translate of a point in  $Q^+$ . On the other hand, any point in  $Q^+$  is a translate of a point in  $\text{Reg}(\tilde{N}^+/N^+)$  by choice of the latter.

The set of points  $\mathcal{A}$  is a subset of the set  $\mathcal{A}^+$  defined in the same way for  $\text{Reg}(\tilde{N}^+/N^+)$  so for the second statement it suffices to prove that  $\mathcal{A}^+$  is finite. For this, note that  $\tilde{N}^+$  is the universal covering of a finite complex, and  $\text{Reg}(\tilde{N}^+/N^+)$  is a subset bounded with respect to any  $\pi_1(N^+)$ -invariant metric. There are only finitely many group elements corresponding to paths of bounded length, so  $\mathcal{A}^+$  is finite. □

By construction we now have a map between fundamental regions denoted by

$$\sigma_{\text{Reg}} : \text{Reg}(\tilde{N}^*/N^*) \rightarrow \text{Reg}(\tilde{D}/D).$$

LEMMA 4.9. *In order to define an equivariant map  $\phi$  on  $\tilde{N}^*$  with finite energy on the fundamental region, it suffices to construct a Lipschitz piecewise  $C^1$  map*

$$\phi_{\text{Reg}} : \text{Reg}(\tilde{N}^*/N^*) \rightarrow \mathcal{T}$$

sending projection-equivalent points to the same point in  $\mathcal{T}$ , and with the property that for any  $x, y \in \text{Reg}(\tilde{N}^*/N^*)$  and any  $\alpha \in \pi_1(N^*)$  such that  $\alpha x = y$ , we should have

$$\phi_{\text{Reg}}(y) = \rho(\alpha)\phi_{\text{Reg}}(x).$$

*Proof.* Suppose that we are given  $\phi_{\text{Reg}}$ . Then define  $\phi : \tilde{N}^* \rightarrow \mathcal{T}$  as follows. For  $u \in \tilde{N}^*$  choose a group element  $g \in \pi_1(N^*)$  such that  $gu \in \text{Reg}(\tilde{N}^*/N^*)$ . Put

$$\phi(u) := \rho(g)^{-1}\phi_{\text{Reg}}(gu).$$

If  $g'$  is another element with  $g'u \in \text{Reg}(\tilde{N}^*/N^*)$ , then putting  $\alpha = g'g^{-1}$ ,  $x = gu$  and  $y = g'u$  we have  $\alpha x = y$  so this is data such as in the hypothesis of the lemma. In this case,

$$\rho(g)^{-1}\phi_{\text{Reg}}(gu) = \rho(g')^{-1}\rho(\alpha)^{-1}\phi_{\text{Reg}}(x) = \rho(g')^{-1}\phi_{\text{Reg}}(\alpha x) = \rho(g')^{-1}\phi_{\text{Reg}}(g'u),$$

which shows that  $\phi$  is well defined. If  $\phi_{\text{Reg}}$  is Lipschitz piecewise  $C^1$ , then the same is true of  $\phi$ . As stated above, by [JZ00], if  $\phi_{\text{Reg}}$  sends projection-equivalent points to the same point in  $\mathcal{T}$ , then it will have finite energy. □

Our strategy will be to use this lemma and try to construct a map which in a certain sense factors through the projection  $\sigma$ . In the most delicate case of unipotent monodromy, this is not exactly possible; however, it will be possible to construct a map which factors over the fundamental region.

According to Lemma 4.4, the centralizer  $Z$  is an invariant. Since the centralizers are different in the three cases of Lemma 4.3, it means that the matrices  $\rho(\gamma_j)$  all fall into the same case of that classification. Thus, for our construction of an initial map we can treat these three cases separately,

as we shall do in the following sections, in increasing order of difficulty: the non-split torus, the split torus, and the unipotent case.

**4.5 The non-split torus case**

In both torus cases, the  $\rho(\gamma_j)$  are commuting matrices of finite order. In particular, there exists a matrix  $A$  which also has finite order, and such that for each  $j$  there is  $m_j$  with  $\rho(\gamma_j) = A^{m_j}$ . In fact,  $A$  can be taken as an appropriate product of powers of the  $\rho(\gamma_j)$ .

Let  $R \subset \mathcal{T}$  be the set of fixed points of  $A$ . The action of  $A$  fixes the ends of  $R$ . This implies that  $A$  is contained in the parabolic subgroup associated to any end. However, since the eigenvalue of  $A$  is not defined over  $K$ , this is impossible: the parabolic subgroups are conjugate to groups of upper-triangular matrices, in particular the eigenvalues of elements of a parabolic subgroup are defined over  $K$ . Thus,  $R$  has no ends, in other words it is compact.

Now, the group  $Z(A)$  fixes  $R$ : if  $\beta \in Z(A)$  and  $x \in R$ , then

$$A(\beta x) = \beta(Ax) = \beta x \implies \beta x \in R.$$

It follows that  $Z(A)$  fixes the barycenter of  $R$ . Let  $R' \subset \mathcal{T}$  be the set of fixed points of  $Z$ . By what we have just said, it is non-empty. Note also that  $A \in Z$  so  $R' \subset R$ , in particular  $R'$  is compact too.

If  $\beta \in NZ$  and  $x \in R'$ , then for any  $\xi \in Z$  we have

$$\xi(\beta x) = \beta(\beta^{-1}\xi\beta x) = \beta x \quad \text{since } \beta^{-1}\xi\beta \in Z,$$

thus  $\beta x \in R'$ . In other words,  $NZ$  sends  $R'$  into itself. Again, this implies that the barycenter of  $R'$  is a fixed point of  $NZ$ .

By Lemma 4.6, the representation  $\rho$  sends  $\pi_1(N^*)$  into  $NZ$ , thus  $\mathcal{T}$  admits a fixed point for the action of  $\pi_1(N^*)$ . We can take as the initial finite-energy map the constant map sending all of  $N^*$  to this fixed point.

**4.6 The split torus case**

As above, there is a matrix  $A \in Z$  of finite order such that the  $\rho(\gamma_j)$  are powers of  $A$ . In the split case,  $A$  may be conjugated to a diagonal matrix with roots of unity on the diagonal. The subset  $R \subset \mathcal{T}$  of points fixed by  $A$  is a real line connecting two ends of  $\mathcal{T}$ . The two ends correspond to the parabolic subgroups of upper and lower triangular matrices. By direct calculation, the centralizer  $Z$ , which is the diagonal torus  $\mathbb{G}_m \subset \text{PSL}(2, K)$ , preserves  $R$ . The normalizer  $NZ$  is the extension of  $Z$  by the Weyl group of transpositions, and again we see directly that  $NZ$  preserves  $R$ , with the non-trivial irreducible component acting by interchanging the two ends, reversing the orientation of  $R$ . By Lemma 4.6, the action of  $\pi_1(N^*)$  via  $\rho$  goes through  $NZ$ , hence it preserves  $R \subset \mathcal{T}$ .

The morphism  $NZ \rightarrow \text{Aut}(R)$  sends  $A$  to the identity. Hence, it sends the normal subgroup generated by the  $\rho(\gamma_j)$  to the identity. This normal subgroup is the kernel of the map  $\pi_1(N^*) \rightarrow \pi_1(N)$ . This map is the same as  $\sigma_* : \pi_1(N^*) \rightarrow \pi_1(D)$ . Therefore, the action of  $\pi_1(N^*)$  on  $R \subset \mathcal{T}$  factors through an action of  $\pi_1(D)$ .

Now  $D$  is a usual compact variety and we can choose an equivariant map  $\tilde{D} \rightarrow R$ . Composing with the projection  $\sigma$  this gives an equivariant map  $\tilde{N}^* \rightarrow R$  which sends projection-equivalent points to the same point in  $R$ , in particular it has finite energy on the fundamental region.

**4.7 The unipotent case**

We would like to mimic the preceding construction. However,  $Z$  is now the unipotent subgroup of upper-triangular matrices with  $\pm 1$  on the diagonal. The fixed-point sets of finitely generated subgroups of  $Z$  are non-empty: they are contractible rays. The action of  $NZ$  will not preserve

these rays. Basically, a diagonal matrix with non-trivial valuations of its diagonal elements will change the fixed ray. If we think of  $Z$  as isomorphic to the additive group of the field  $K$ , then any finitely generated subgroup goes into a subgroup of the form  $t^q \mathcal{O}_K$  where  $t$  is the uniformizing parameter. The ray of fixed points depends on  $q$  and conjugating with a diagonal matrix can change  $q$ .

We get around this problem by noting that we really only need to choose a map on the fundamental region, equivariant with respect to the set of group elements which serve to identify different points of this region. This set is finite, so we can start far enough out along the fixed ray so that the finite set of translations involved does not take us out of the set of fixed points. We obtain a map from  $\text{Reg}(\tilde{N}^*/N^*)$  to  $\mathcal{T}$  which factors through  $\text{Reg}(\tilde{D}/D)$ , even though the resulting equivariant map on the whole universal covering  $\tilde{N}^*$  will not factor through  $\tilde{D}$ . The different copies of  $\text{Reg}(\tilde{N}^*/N^*)$  which make up  $\tilde{N}^*$  are related by paths which, even if they project to the same path in  $\pi_1(D)$ , act differently on  $\mathcal{T}$  and this serves to create a divergence in the equivariant map.

We now explain the argument more carefully.

**COROLLARY 4.10.** *Suppose that we can construct a Lipschitz piecewise  $C^1$  map  $\varphi : \text{Reg}(\tilde{D}/D) \rightarrow \mathcal{T}$  with the property that for any  $x, y \in \text{Reg}(\tilde{N}^*/N^*)$  and any  $\alpha \in \pi_1(N^*/N)$  with  $\alpha x = y$ , then*

$$\varphi(\sigma_{\text{Reg}} y) = \rho(\alpha)\varphi(\sigma_{\text{Reg}} x).$$

*With this, posing  $\phi_{\text{Reg}}(y) := \varphi(\sigma_{\text{Reg}}(y))$  leads to an equivariant map  $\phi : \tilde{N}^* \rightarrow \mathcal{T}$  which has finite energy on the fundamental region.*

*Proof.* Use Lemma 4.9. The criterion of that lemma is exactly the same as the hypothesis here. Note that if  $\phi_{\text{Reg}}$ , factors through  $\sigma_{\text{Reg}}$ , then it automatically sends projection-equivalent points to the same point, since  $\sigma$  has this property. □

Define the subsets of group elements corresponding to boundary identifications of the fundamental regions:

$$\mathcal{A} := \{\alpha \in \pi_1(N^*), \exists x, y \in \text{Reg}(\tilde{N}^*/N^*), \alpha x = y\},$$

and

$$\mathcal{B} := \{\beta \in \pi_1(D), \exists x, y \in \text{Reg}(\tilde{D}/D), \beta x = y\}.$$

Both of these subsets are finite, and we have a map  $\mathcal{A} \rightarrow \mathcal{B}$ .

Let  $\mathcal{A} \times_{\pi_1(D)} \mathcal{A}$  be the set of pairs  $(\alpha, \beta) \in \mathcal{A} \times \mathcal{A}$  with  $\sigma_*(\alpha) = \sigma_*(\beta)$  in  $\pi_1(D)$ .

**LEMMA 4.11.** *The map  $\mathcal{A} \rightarrow \mathcal{B}$  is surjective, in other words  $\mathcal{B}$  is the image of  $\mathcal{A}$  in  $\pi_1(D)$  and it is the quotient of  $\mathcal{A}$  by the equivalence relation  $\mathcal{A} \times_{\pi_1(D)} \mathcal{A}$ .*

*Proof.* Look at the choice of  $\text{Reg}(\tilde{N}^*/N^*)$  described above Lemma 4.8. We first claim that the map  $\sigma_{\text{Reg}} : \text{Reg}(\tilde{N}^*/N^*) \rightarrow \text{Reg}(\tilde{D}/D)$  is surjective. To see this, note that  $N^* \rightarrow D$  is surjective, so

$$N^* \times_D \text{Reg}(\tilde{D}/D) \rightarrow \text{Reg}(\tilde{D}/D)$$

is surjective. On the other hand, any point  $N^* \times_D \text{Reg}(\tilde{D}/D)$  lifts to a point of  $\text{Reg}(\tilde{N}^*/N^*)$  by construction of this latter. This proves the claim.

Now we continue with the proof of the lemma. Suppose that  $\beta \in \mathcal{B}$  with  $u, v \in \text{Reg}(\tilde{D}/D)$  such that  $\beta u = v$ . Choose  $x \in \text{Reg}(\tilde{N}^*/N^*)$  such that  $\sigma_{\text{Reg}}(x) = u$ . Choose  $\beta' \in \pi_1(N^*)$  mapping to  $\beta$ , and put  $y' := \beta'x$ . We have  $\tilde{\sigma}(y') = \beta u = v$ . Thus,

$$y' \in \tilde{\sigma}^{-1}(\text{Reg}(\tilde{D}/D)) = Q^*.$$

By the choice of  $\text{Reg}(\tilde{N}^*/N^*)$  as a fundamental region for the action of  $\ker(\sigma_*)$  on  $Q^*$ , there is a group element  $g \in \ker(\sigma_*)$  such that  $y := g \cdot y' \in \text{Reg}(\tilde{N}^*/N^*)$ . Now, putting  $\alpha := g\beta'$  we obtain  $\alpha x = g\beta'x = gy' = y$ , so  $\alpha \in \mathcal{A}$  and it maps to  $\beta$ . □

LEMMA 4.12. *Suppose that we can construct a map  $\varphi : \text{Reg}(\tilde{D}/D) \rightarrow \mathcal{T}$  with the property that for any  $u, v \in \text{Reg}(\tilde{D}/D)$  and any  $\alpha \in \mathcal{A}$  such that  $\sigma_*(\alpha)u = v$ , then  $\varphi(v) = \rho(\alpha)\varphi(u)$ . Then  $\varphi$  satisfies the criterion of Corollary 4.10.*

*Proof.* Suppose that  $x, y \in \text{Reg}(\tilde{N}^*/N^*)$  and  $\alpha \in \pi_1(N^*/N)$  with  $\alpha x = y$ . Note that, by definition,  $\alpha \in \mathcal{A}$ . Put  $u := \sigma_{\text{Reg}}(x)$  and  $v := \sigma_{\text{Reg}}(y)$ . We have  $\sigma_*(\alpha)u = v$ . By the hypothesis of this lemma,  $\varphi(v) = \rho(\alpha)\varphi(u)$ , and in view of the definition of  $u$  and  $v$  that gives exactly the condition of Corollary 4.10. □

Note that  $D$  is obtained from  $\text{Reg}(\tilde{D}/D)$  by glueing together points  $u, v$  whenever there is  $\beta \in \mathcal{B}$  such that  $\beta u = v$ .

COROLLARY 4.13. *In order to apply Corollary 4.10 to construct a finite-energy initial map, it suffices to construct a map  $\varphi : \text{Reg}(\tilde{D}/D) \rightarrow \mathcal{T}$  with the property that for any two points  $u, v \in \text{Reg}(\tilde{D}/D)$  which are equivalent by a group element  $\beta \in \mathcal{B} \subset \pi_1(D)$ , then for any lifting  $\alpha \in \mathcal{A}$  with  $\sigma_*(\alpha) = \beta$  we should have  $\varphi(v) = \rho(\alpha)\varphi(u)$ .*

*Proof.* Given that the condition here makes sense by Lemma 4.11, the statement is the same as that of Lemma 4.12. □

We have to look carefully at the standard procedure for creating an initial equivariant map into a contractible space, because  $\pi_1(D)$  does not actually act naturally on  $\mathcal{T}$ .

PROPOSITION 4.14. *Suppose that we have a nested sequence of subsets*

$$R_{-1} \subset R_0 \subset R_1 \subset \dots \subset R_d \subset \mathcal{T}$$

where  $d = \dim_{\mathbb{R}}(D)$ , such that each  $R_k$  is contractible, such that  $\rho(\mathcal{A} \times R_k) \subset R_{k+1}$  and such that

$$\rho(\alpha)t = \rho(\beta)t, \quad \text{for all } t \in R_k \text{ and all } (\alpha, \beta) \in \mathcal{A} \times_{\pi_1(D)} \mathcal{A}.$$

Then we can construct an initial map  $\varphi$  as required by the previous Corollary 4.13.

*Proof.* Let  $\text{Reg}_k(\tilde{D}/D)$  be the  $k$ -skeleton of the fundamental region. Define inductively maps  $\varphi_k : \text{Reg}_k(\tilde{D}/D) \rightarrow R_k$  as follows. For  $k = -1$  the domain is empty and the map is trivially defined. For any  $k$ , assuming that the map is defined on the  $k - 1$ -skeleton, then choose a subset of  $k$ -simplices of  $\text{Reg}_k(\tilde{D}/D)$  mapping isomorphically to the set of  $k$ -simplices of  $D$ . By contractibility of  $R_{k-1}$  we can map these simplices into  $R_{k-1}$  arbitrarily, given the map  $\varphi_{k-1}$  which we already know on their boundaries. The remaining  $k$ -simplices of  $\text{Reg}_k(\tilde{D}/D)$  are mapped into  $\mathcal{T}$  in a unique and well-defined way using the group translation condition for  $\varphi$ . Note that any  $k$ -simplex of  $\text{Reg}_k(\tilde{D}/D)$  is related to one in the original set by a unique group element because the group  $\pi_1(D)$  acts freely on  $\tilde{D}$ , and furthermore this group element is in  $\mathcal{B}$  by the definition of the latter. It lifts to an element of  $\mathcal{A}$  by Lemma 4.11. Thus, with our condition that  $\rho(\mathcal{A} \times R_{k-1}) \subset R_k$  we get that these maps send all of  $\text{Reg}_k(\tilde{D}/D)$  into  $R_k$ . This completes the inductive construction of  $\varphi$ . □

We have now reduced to a purely algebraic problem. Recall that the centralizer  $Z$  is the subgroup of unipotent upper-triangular matrices, its normalizer is  $NZ = P$  the group of upper-triangular matrices in  $\text{PSL}(2, K)$ , and  $NZ/Z \cong \mathbb{G}_m$  is the diagonal torus.

LEMMA 4.15. *Let  $P$  be the group of upper-triangular matrices in  $\text{PSL}(2, K)$ , with projection  $\xi : P \rightarrow \mathbb{G}_m$  to the diagonal torus. Suppose that we have a finite subset  $A \subset P$ . For any  $d$  there is a sequence of contractible real rays  $R_{-1} \subset \dots \subset R_d \subset \mathcal{T}$ , such that  $A \cdot R_k \subset R_{k+1}$  and*

$$\text{for all } \alpha, \beta \in A, \quad \xi(\alpha) = \xi(\beta) \quad \Rightarrow \quad \text{for all } t \in R_k, \quad \alpha \cdot t = \beta \cdot t.$$

*Proof.* Let

$$D := \{(\alpha\beta^{-1}) \in Z, \text{ for } \alpha, \beta \in A \text{ with } \xi(\alpha) = \xi(\beta)\}.$$

This is a finite subset of  $Z$ . We can rewrite the conditions of the lemma by saying that we want rays  $R_k$  for  $-1 \leq k \leq d$  such that  $R_k$  are fixed by elements of  $D$ , and  $A \cdot R_k \subset R_{k+1}$ .

If we consider the standard basis  $\{e_1, e_2\}$  for  $K^2$ , it corresponds to two ends of  $\mathcal{T}$  which are the two parabolic subgroups of upper- and lower-triangular matrices. The real line joining these two ends consists of the segments joining together all adjacent lattices of the form  $M_q := \langle t^q e_1, e_2 \rangle$  where  $t \in K$  is the uniformizing parameter. The group of points of  $Z$  is just  $K$ , and an element  $a \in K$  acts on the basis elements by  $e_1 \mapsto e_1, e_2 \mapsto ae_1 + e_2$ . There is an upper bound on the valuation of elements of the finite subset  $D \subset K$ . This leads to  $q'$  such that the lattice  $M_q$  is fixed by  $D$  for any  $q \leq q'$ . We may consider any  $q_k \leq q_0$  and define  $R_k$  to be the ray of segments joining the lattices  $M_q$  for  $q \leq q_k$ . With this choice we obtain the first condition that  $R_k$  consists of fixed points of  $D$ .

To obtain the second condition, note that the diagonal group  $\mathbb{G}_m \cong K^*$  acts on our real line by translations through the valuation map  $\nu : K^* \rightarrow \mathbb{Z}$ . An element  $b \in K^*$  sends  $M_q$  to  $M_{q+\nu(b)}$ . Again, since  $\xi(A) \subset K^*$  is a finite subset, there is a bound on the valuations of elements. Thus, there is  $n$  such that  $\nu(b) \leq n$  for  $b \in \xi(A)$ . If we choose  $q_{k+1} = q_k + n$ , for example, then the second condition  $A \cdot R_k \subset R_{k+1}$  will hold.

To complete the proof, just choose  $q_{-1} \leq q' - (d + 1)n$ , then  $q_k = q_{-1} + (k + 1)n \leq q'$  for every  $k = -1, \dots, d$  and we obtain the required choice of rays. □

This suffices for our construction of  $\varphi$ .

**COROLLARY 4.16.** *In the unipotent case, putting  $A := \rho(\mathcal{A}) \subset P \subset \text{PSL}(2, K)$  and  $R_k := R^{d-k}$  we obtain regions satisfying the criterion of Proposition 4.14, hence going back to Corollary 4.10 we obtain the desired finite-energy initial map.*

*Proof.* By the map of exact sequences in Lemma 4.6, the points of  $\mathcal{A} \times_{\pi_1(D)} \mathcal{A}$  map to  $A \times_{\mathbb{G}_m} A$ . Thus, the subsets provided by Lemma 4.15 serve to satisfy the criterion of Proposition 4.14. □

*Remark 2.* In the above construction for the unipotent case, we could have replaced the divisor  $D$  by its incidence complex  $\Sigma$ , cf. Remark 1. There is a projection  $D \rightarrow \Sigma$  and composing gives a projection  $N^* \rightarrow \Sigma$ . We could have constructed the initial map in a way which factors through this projection over the fundamental region. This is special to the unipotent case: in the split torus case it looks like we need to consider the projection to  $D$ .

### 4.8 The pluriharmonic map

We sum up the above constructions and apply Gromov–Schoen theory to deduce the existence of an equivariant harmonic map of finite energy on the fundamental region. In this section, we no longer assume that  $D$  is connected.

Our now arbitrary basepoint  $x \in X$  corresponds to a universal covering  $f : \tilde{X} \rightarrow X$ .

Let  $D_{(i)}$  denote a connected component of the divisor  $D$ . It may be a union of several irreducible components. Let  $N_{(i)}$  denote a tubular neighborhood of  $D_{(i)}$  in  $\overline{X}$  and let  $N_{(i)}^* := N_{(i)} \cap X$  denote the ‘punctured’ tubular neighborhood. Define the fundamental group of  $N_{(i)}^*$  inside  $\pi_1(X, x)$  by choosing a basepoint  $x_{(i)}$  in  $N_{(i)}^*$  and joining it to  $x$  by a path. Denote the subgroup of paths which consist of going from  $x$  to  $x_{(i)}$  along our path, then arbitrarily inside  $N_{(i)}^*$ , and again along our path back to  $x$ , by  $\pi_1(N_{(i)}^*) \subset \pi_1(X, x)$ .

Let  $\tilde{N}_{(i)}^*$  denote the universal covering determined by the basepoint  $x_{(i)}$ . The inverse image  $f^{-1}N_{(i)}^*$  inside  $\tilde{X}$  is a disjoint union of copies of  $\tilde{N}_{(i)}^*$ . In order to define a  $\pi_1(X, x)$ -equivariant map

starting from this disjoint union, it is equivalent to defining a  $\pi_1(N_{(i)}^*, x_{(i)})$ -equivariant map starting from  $\tilde{N}_{(i)}^*$ .

Our preceding discussion where we focalized on a single connected component of  $D$ , allows us to choose a  $\pi_1(N_{(i)}^*, x_{(i)})$ -equivariant map  $\tilde{N}_{(i)}^* \rightarrow \mathcal{T}$  which has finite energy on a fundamental region. This gives a  $\pi_1(X, x)$ -equivariant map  $\tilde{X} \supset f^{-1}N_{(i)}^* \rightarrow \mathcal{T}$ . Do this for each connected component  $D_{(i)}$ . By contractibility of  $\mathcal{T}$  we can extend these maps to a  $\pi_1(X, x)$ -equivariant map  $\phi : \tilde{X} \rightarrow \mathcal{T}$ . The procedure was recalled in detail in the proof of Proposition 4.14. The complement of the union of translates of the fundamental regions for the  $\tilde{N}_{(i)}^*$  is relatively compact in the fundamental region of  $\tilde{X}$ , so the map  $\phi$  has finite energy on the fundamental region.

The action of  $\pi_1(X)$  on  $\mathcal{T}$  is reductive in the sense of [JZ00], because of our hypothesis that the Zariski closure of the image of  $\rho$  is  $SL(2, K)$ .

By the general theory of Gromov–Schoen [GS92] our initial map can be replaced by a harmonic  $\rho$ -equivariant map  $\Phi : \tilde{X} \rightarrow \mathcal{T}$ .

The harmonic map is, in fact, pluriharmonic, by [GS92, Theorem 7.3].

The hypothesis that the image of  $\rho$  is not contained in a compact subgroup, implies that the pluriharmonic map is not constant, because if it were constant, then its image point would be fixed by  $\rho$  and the subgroup of elements fixing a point of  $\mathcal{T}$  is compact. We obtain the following result.

**PROPOSITION 4.17.** *Under Hypothesis 4.1, there exists a non-constant pluriharmonic  $\rho$ -equivariant map from the universal cover of  $X$  to the Bruhat–Tits tree  $\mathcal{T}$ . With respect to the Poincaré-like complete finite-volume metric on  $X$ , the pluriharmonic map has finite energy.*

### 4.9 Fixed points of normal subgroups

We close out this section with a study of the following situation. Suppose that a finitely presented group  $\Gamma$  acts on  $\mathcal{T}$  by a representation  $\rho : \Gamma \rightarrow PSL(2, K)$ , and suppose that  $\Upsilon \subset \Gamma$  is a normal subgroup. Let  $R \subset \mathcal{T}$  be the set of fixed points of  $\Upsilon$ , that is the set of points  $y \in \mathcal{T}$  such that  $uy = y$  for all  $u \in \Upsilon$ . Note that  $R$  is convex in the sense that if  $y, z \in R$ , then the unique path joining  $y$  to  $z$  is contained in  $R$ . Thus,  $R$  is a subtree of  $\mathcal{T}$ .

The fact that  $\Upsilon$  is a normal subgroup implies that  $\Gamma R \subset R$ , indeed

$$\text{for all } u \in \Upsilon, \text{ all } \gamma \in \Gamma \text{ for all } y \in R, \quad u(\gamma y) = \gamma((\gamma^{-1}u\gamma)y) = \gamma y.$$

In particular, we obtain an action of the quotient  $\Gamma/\Upsilon$  on  $R$ .

The subgroup  $\Upsilon$  fixes the set of ends of  $R$  (which are also ends of  $\mathcal{T}$ ).

If  $R$  has at least three distinct ends, this implies that the image of  $\Upsilon$  in  $\mathcal{T}$  is contained in the intersection of at least three distinct normal subgroups. These normal subgroups correspond to distinct one-dimensional subspaces  $L_i \subset K^2$  preserved by  $\Upsilon$ . Up to conjugation we may assume that three of these subspaces are the two coordinate lines plus the diagonal. The subgroup of elements preserving these subspaces is the center  $\{\pm 1\} \subset SL(2, K)$ , that is the trivial subgroup of  $PSL(2, K)$ . Thus, if  $R$  has at least three ends, then the image of  $\Upsilon$  is trivial.

On the other hand, the set of ends of  $R$  remains invariant, although it may be acted upon, by the action of  $\Gamma$ . A subset consisting of one or two ends corresponds to a subgroup (either a parabolic subgroup or a torus), so in either of these cases the image of  $\Gamma$  cannot be Zariski-dense in  $PSL(2, K)$ .

Finally, if  $R$  has no ends, thus it is compact, then  $\Gamma$  fixes its barycenter so  $\Gamma$  is contained in a compact subgroup. We can sum this up in the following lemma.

**LEMMA 4.18.** *Suppose that  $\rho : \Gamma \rightarrow PSL(2, K)$  is a representation, and suppose that  $\Upsilon \subset \Gamma$  is a normal subgroup acting with a non-empty set of fixed points. Suppose that the image  $\rho(\Gamma)$*



is Zariski-dense and not contained in a compact subgroup of  $\mathrm{PSL}(2, K)$ . Then the image of  $\Upsilon$  is trivial.

It will also be useful to have the invariance of our hypothesis under generically finite coverings.

LEMMA 4.19. *Suppose that  $(X, \overline{X}, \rho)$  satisfies Hypothesis 4.1. Then for any generically surjective map  $Z \rightarrow X$  and any smooth normal crossings compactification  $Z \subset \overline{Z}$ , if we let  $\rho_Z$  denote the pullback of  $\rho$  to  $Z$  and  $Y$  the maximal open subset of  $\overline{Z}$  on which  $\rho_Z$  extends to  $\rho_Y$ , then  $(Y, \overline{Z}, \rho_Y)$  satisfies Hypothesis 4.1.*

*Proof.* Note first that  $Y$  is still the complement of a divisor with normal crossings, which is the divisor  $\overline{Z} - Z$  with possibly some components removed.

The local monodromy transformations on  $Y$  come from those of  $X$ . If we have a small loop around a component of the divisor at infinity in  $Y$ , then it goes to a loop in one of the neighborhoods of a multiple intersection point on  $D$ . In such a neighborhood, the monodromy is an abelian representation whose generators are quasiunipotent. A commuting family of quasiunipotent matrices can be simultaneously upper-triangularized, so any matrix in the group they generate is also quasiunipotent. This shows that the monodromy of  $\rho_Y$  is quasiunipotent at infinity.

The image of  $\pi_1(Z)$  is of finite index in  $\pi_1(X)$ , call this subgroup  $\Upsilon$ . Furthermore,  $\rho_Y(\pi_1(Y)) = \rho_Z(\pi_1(Z)) = \rho(\Upsilon)$ . Let  $\Upsilon'$  be the intersection of the conjugates of  $\Upsilon$ . It is a normal subgroup. Let  $H$  be the Zariski closure of  $\rho(\Upsilon')$ . It is a normal subgroup of  $\mathrm{SL}(2, K)$  since the image of  $\rho$  is Zariski-dense, so  $H$  is either the whole group or the center. If it were the center, then the image of  $\pi_1(X)$  in  $\mathrm{PSL}(2, K)$  would be finite, contradicting the Zariski-density of  $\rho$ . Thus,  $H = \mathrm{PSL}(2, K)$ , which shows that  $\rho_Y$  is Zariski-dense.

Suppose now that  $\rho_Y(\pi_1(Y))$  were contained in a compact subgroup. Using the same group  $\Upsilon'$  as in the previous paragraph, we would obtain  $\rho(\Upsilon')$  contained in a compact subgroup. By Lemma 4.18 the image  $\rho(\Upsilon')$  would have to be contained in the center, again contradicting Zariski-density of  $\rho$ . This completes the verification of the conditions of Hypothesis 4.1.  $\square$

## 5. Properties of the pluriharmonic map

In this section, we look further at the properties of the pluriharmonic mapping constructed above. The case of trees was discussed explicitly in [GS92], when  $X$  is compact. For the quasiprojective case, our situation is subsumed under the more general case of maps to buildings considered by Jost and Zuo [JZ00]. Since the case of maps to trees is considerably simpler in that there is only a single spectral form to consider, it seems opportune to look at the details here.

We keep Hypothesis 4.1 and look at the pluriharmonic map given by Proposition 4.17. Our goal is to obtain the factorization result Theorem 5.13 given at the end.

### 5.1 The spectral form

The pluriharmonic map yields spectral data which is a two-valued one-form  $\theta$  on  $X$ . This is easy to describe in the case of a map to a tree: the edges of  $\mathcal{T}$  have a linear structure and also metrics so we can identify them with intervals in  $\mathbb{R}$ . This identification is well defined up to translation and inversion, so the differential  $d\theta$  is well defined up to  $\pm 1$ . The differential is pluriharmonic, so it is the real part of a holomorphic differential form  $\theta$ . It turns out, due to the regularity statements of Gromov and Schoen [GS92] which are local so they work equally well in the quasiprojective case, that the symmetric square  $\mathrm{Sym}^2(\theta)$  is a well-defined and holomorphic section of  $\mathrm{Sym}^2(\Omega_X^1)$ .

Near a smooth point of the divisor at infinity, choose a coordinate system  $(z_1, \dots, z_n)$  such that  $D$  is given by  $z_1 = 0$ . Then we can write

$$\text{Sym}^2(\theta) = \sum_{i \leq j} a_{ij}(z) dz_i dz_j$$

with the functions  $a_{ij}(z)$  holomorphic on the complement of  $z_1 = 0$ .

At any point  $z \in X$ ,  $\text{Sym}^2(\theta)$  is the square of a one-form  $\theta(z)$  well defined up to sign. Furthermore, the energy is the integral of the square norm of  $\theta(z)$ , which in turn is the same as  $|\text{Sym}^2(\theta)|$ . We conclude that the integral (over our coordinate neighborhood) is finite:

$$\int \left| \sum_{i \leq j} a_{ij}(z) dz_i dz_j \right| d\mu < \infty.$$

Up to a bounded change in the metric on the base, we may assume that the metric is the product of the Poincaré metric in  $z_1$  with the euclidean metric on the  $z_i$  for  $i > 1$ . In particular, the  $dz_i dz_j$  are orthogonal and we obtain

$$\int |a_{ij}(z) dz_i dz_j| d\mu < \infty$$

for all  $i \leq j$ . Recall that  $|dz_i| = 1$  for  $i > 1$  and if we set  $r := |z_1|$

$$|dz_1| = r|\log r|$$

and

$$d\mu = \frac{dz_1 \wedge \bar{d}z_1 \wedge \dots \wedge dz_n \wedge \bar{d}z_n}{r^2 |\log r|^2}.$$

To show the meromorphy of  $a_{ij}$  along  $D$  and to calculate their maximum order of poles, it suffices to restrict to curves given by  $z_i = p_i$  fixed for  $i > 1$ . This is because if the integral is finite, then it is finite for almost all curves, and we have a Hartogs principle for separately almost everywhere meromorphic functions [Shi86, Shi94, Sic69]. The slices in the other directions are automatically almost everywhere holomorphic, indeed for any  $z_1 \neq 0$ .

Thus, we are reduced to considering a function  $a_{ij}^p(z_1) := a_{ij}(z_1, p_2, \dots, p_n)$  of one variable, holomorphic on a punctured neighborhood of the origin, with the hypothesis that

$$\int |a_{ij}^p(u)| (r|\log r|)^{m-2} du \bar{d}u < \infty,$$

where  $m = 2$  if  $i = j = 1$  or  $m = 1$  if  $i = 1 < j$  and  $m = 0$  if  $i, j > 1$ . Here  $r = |u|$ .

We are in a situation similar to Bell’s removable singularities theorem [Bel82, p. 687].

An argument using Cauchy’s theorem on disks not meeting the origin shows that  $a_{ij}^p$  is meromorphic at the origin. To obtain the maximum order of poles, suppose that

$$a_{ij}^p = bu^{-n} + \dots$$

is the Laurent expansion with first coefficient  $b \neq 0$ . Then for  $u$  near the origin

$$|a_{ij}^p(u)| \sim r^{-n},$$

so we have

$$\int_0^1 r^{m-n-1} |\log r|^{m-2} dr < \infty.$$

Making the change of variables  $u = -\log r$  we obtain

$$\int_0^\infty e^{(n-m)u} u^{m-2} du < \infty.$$

We conclude that in the case  $m = 0$  we obtain  $n \leq 0$ , whereas for  $m \geq 1$  we obtain  $n < m$ . Thus, for  $m = 0, 1$  we have  $n \leq 0$  and for  $m = 2$  we have  $n \leq 1$ . In view of the definition of  $m$  this means that the  $a_{ij}^p$  are holomorphic on  $\overline{X}$  if either of  $i$  or  $j$  is different from one, whereas  $a_{11}^p$  has a pole of order at most one.

Going back to all of  $\overline{X}$ , we conclude by Hartogs principle [Shi86, Shi94, Sic69] that the  $a_{ij}$  are meromorphic along smooth components of the divisor. By the more classical Hartogs principle, they are meromorphic everywhere. Their orders of poles are bounded as described previously.

We can say what this bound means in invariant terms. Define the *half-logarithmic symmetric square* to be the subsheaf denoted, for lack of better notation, by

$$\text{Sym}^2(\Omega_{\overline{X}}^1; \frac{1}{2} \log D) \subset j_*(\text{Sym}^2(\Omega_X^1))$$

as the subsheaf which is generated along smooth points of  $D_i$  by  $\text{Sym}^2(\Omega_{\overline{X}}^1)$  plus  $dz^2/z$  where  $z$  is a coordinate of  $D_i$ . One can check that the subsheaf  $\text{Sym}^2(\Omega_{\overline{X}}^1; \frac{1}{2} \log D)$  is independent of the choice of coordinate. This defines the subsheaf at points of  $\overline{X}$  which are not normal crossing points of  $D$ . This defines a unique reflexive coherent sheaf on  $\overline{X}$ . By choosing local coordinates at a normal crossing point and noting that the generators of  $\mathcal{F}$  have a simple monomial form, we can see that  $\mathcal{F}$  is actually locally free.

Our estimate can be stated as follows.

LEMMA 5.1. *With the above notation, for the finite energy pluriharmonic map the differential  $\text{Sym}^2(\theta)$  extends to a holomorphic section of the bundle  $\text{Sym}^2(\Omega_{\overline{X}}^1; \frac{1}{2} \log D)$  over  $\overline{X}$  defined in the previous paragraph.*

LEMMA 5.2. *Suppose that  $p : \overline{Z} \rightarrow \overline{X}$  is a morphism from a smooth projective variety to  $\overline{X}$ , and suppose that we have a divisor  $D_Z \subset \overline{Z}$  which is assumed to have normal crossings. Suppose that the inverse image of  $D$  is contained in  $D_Z$ . Thus, if  $Z := \overline{Z} - D_Z$ , then  $p : Z \rightarrow X$ . Define as above the half-logarithmic symmetric squares for  $(\overline{X}, D)$  and  $(\overline{Z}, D_Z)$ . Then the natural map*

$$p^*\text{Sym}^2(\Omega_X^1) \rightarrow \text{Sym}^2(\Omega_Z^1)$$

*extends to a map*

$$p^*\text{Sym}^2(\Omega_{\overline{X}}^1; \frac{1}{2} \log D) \rightarrow \text{Sym}^2(\Omega_{\overline{Z}}^1; \frac{1}{2} \log D_Z).$$

*Proof.* The sheaves on both sides of the desired map are vector bundles. Thus, to obtain the map we just need to construct it in codimension one. In particular, we can look near to a point on a smooth point  $P \in D_Z$ , where we have coordinates  $z_1, \dots, z_n$  such that the divisor  $D_Z$  is given by  $z_1 = 0$  (with  $P$  corresponding to  $z_i = 0$ ). The image  $Q = p(P)$  in  $X$  is a point which we may assume is on  $D$ , because if it were not then the statement would be immediate. Thus, we have coordinates  $x_1, \dots, x_m$  on  $X$  again centered at  $Q$  and with  $D$  given by  $x_1 \cdots x_k = 0$ . Write

$$x_i \circ p = u_i(z)z_1^{a_i},$$

with  $u_i$  non-vanishing along  $D_Z$  at the point  $P$ . The inverse image of  $D$  is given by the equation

$$0 = (x_1 \cdots x_k) \circ p = u_1(z) \cdots u_k(z)z_1^{a_1 + \cdots + a_k}.$$

By hypothesis, the zero set of this equation is contained in  $D_Z$ , which near  $P$  is just the smooth divisor  $z_1 = 0$ . This implies that the zero sets of the  $u_i(z)$  do not contain any other irreducible components for  $i = 1, \dots, k$ . Therefore,  $u_i(P) \neq 0$  for  $i = 1, \dots, k$ . This implies that  $a_i > 0$  for  $i = 1, \dots, k$  because our coordinates  $x_i$  are centered at  $Q$  so  $x_i \circ p(P) = 0$ . Of course, we cannot say anything for the  $u_{k+1} \cdots u_n$  but these will not enter into the subsequent argument.

Now

$$\text{Sym}^2(\Omega_{\overline{X}}^1; \frac{1}{2} \log D)$$

is locally free, with generators near  $Q$  of the form

$$\begin{aligned} \frac{dx_i^2}{x_i}, & \quad 1 \leq i \leq k, \\ dx_j^2, & \quad j > k, \end{aligned}$$

and

$$dx_i dx_j, \quad 1 \leq i < j \leq n.$$

Note that the pullback of  $dx_i$  is

$$p^* dx_i = a_i u_i(z) z_1^{a_i-1} dz_1 + \sum \frac{\partial u_i}{\partial z_j} z_1^{a_i} dz_j.$$

This is a section of  $\Omega_{\mathbb{Z}}^1$ , so the pullbacks of the generators  $dx_j^2$  and  $dx_i dx_j$  are in  $\text{Sym}^2 \Omega_{\mathbb{Z}}^1$ . We just have to check the pullback of the fractional generator for  $i = 1, \dots, k$ . We have

$$p^* \left( \frac{dx_i^2}{x_i} \right) = z_1^{2a_i-3} a_i^2 u_i(z)^2 dz_1^2 + z_1^{2a_i-2} A$$

where  $A$  is a section of  $\text{Sym}^2 \Omega_{\mathbb{Z}}^1$ . The fact that  $a_i \geq 1$  means that the term with  $A$  is a section of  $\text{Sym}^2 \Omega_{\mathbb{Z}}^1$ . Similarly, the first term is a holomorphic multiple of  $dz_1^2/z_1$ . Thus,  $p^*(dx_i^2/x_i)$  is a section of  $\text{Sym}^2(\Omega_{\mathbb{Z}}^1; \frac{1}{2} \log D_Z)$  as required. □

**COROLLARY 5.3.** *Suppose that  $\Phi$  is a finite-energy pluriharmonic map on  $X$  corresponding to a representation satisfying Hypothesis 4.1. Suppose that  $p : \mathbb{Z} \rightarrow \overline{X}$  is as in Lemma 5.2. Then the pullback  $\Phi \circ p$  is a pluriharmonic map on  $Z$  satisfying the condition of Lemma 5.1, that the symmetric square of its differential is a section of  $\text{Sym}^2(\Omega_{\mathbb{Z}}^1; \frac{1}{2} \log D_Z)$ .*

*Proof.* Combine Lemmas 5.1 and 5.2. □

### 5.2 Uniqueness

The pluriharmonic map itself may not be unique. However, its differential is unique, and different pluriharmonic maps are related by a geodesic homotopy which preserves the differential.

Suppose that  $\Phi_0$  and  $\Phi_1$  are two different pluriharmonic maps of finite energy equivariant for  $\rho$ . Define the *geodesic homotopy*  $\{\Phi_t\}_{t \in [0,1]}$  as follows. For any point  $x \in \tilde{X}$ , let  $\Phi_t(x)$  denote the geodesic path joining  $\Phi_0(x)$  and  $\Phi_1(x)$ .

**LEMMA 5.4.** *The geodesic homotopy is a continuous, piecewise differentiable map*

$$\Phi : [0, 1] \times \tilde{X} \rightarrow \mathcal{T}.$$

*There is a dense open set  $U \in \tilde{X}$  such that on  $[0, 1] \times U$  the map  $\Phi$  is harmonic in the second variable and linear in the first variable. On each connected component of  $U$  there is a choice of chart in  $\mathcal{T}$  containing the images of all  $\Phi_t(x)$ , and in this chart the differential with respect to the second variable satisfies*

$$d\Phi_t(x) = t d\Phi_1(x) + (1 - t) d\Phi_0(x).$$

*Proof.* Consider the function  $G(x, y, t)$  which to points  $x, y \in \mathcal{T}$  and  $t \in [0, 1]$  associates the point in  $\mathcal{T}$  on the geodesic from  $x$  to  $y$ , with coordinate  $t$ . Thus,  $G(x, y, 0) = x$  and  $G(x, y, 1) = y$ . A geometric picture shows that it is Lipschitz, hence continuous. Furthermore, if  $x$  and  $y$  are not vertices of  $\mathcal{T}$  then near  $(x, y)$  the function is bilinear onto the flat of  $\mathcal{T}$  joining  $x$  to  $y$ . In terms of a real coordinate  $r$  on this flat,  $G$  has the form

$$rG(x, y, t) = (1 - t)r(x) + tr(y).$$

Our function  $\Phi$  is defined by

$$\Phi(t, x) := G(\Phi_0(x), \Phi_1(x), t).$$

The open set  $U$  is the set of points  $x$  such that  $\Phi_0(x)$  and  $\Phi_1(x)$  are not vertices of  $\mathcal{T}$ . This is the complement of a one-dimensional real analytic subvariety, in particular it is dense. On  $U$ , the map  $\Phi$  is linear in  $t$  and harmonic in  $x$  because it is a linear combination of harmonic functions. In terms of a real coordinate  $r$  on the flat joining  $\Phi_0(x)$  to  $\Phi_1(x)$ , we have

$$r\Phi_t(x) = (1 - t)r\Phi_0(x) + t r\Phi_1(x).$$

This gives the formula for the differential. □

COROLLARY 5.5. *The energy of  $\Phi_t$  satisfies*

$$\int_X |d\Phi_t|^2 = e_0 + e_1 t + e_2 t^2$$

with

$$e_2 = \int_X |d\Phi_0 - d\Phi_1|^2.$$

Here the  $d\Phi_0$  and  $d\Phi_1$  are taken using the charts of  $\mathcal{T}$  as in the previous lemma, well defined on connected components of  $U$ .

*Proof.* Plug in the formula

$$d\Phi_t(x) = t d\Phi_1(x) + (1 - t) d\Phi_0(x)$$

to obtain

$$\begin{aligned} |d\Phi_t|^2 &= |t d\Phi_1(x) + (1 - t) d\Phi_0(x)|^2 \\ &= t^2 |d\Phi_1(x)|^2 + (1 - t)^2 |d\Phi_0(x)|^2 + 2t(1 - t) d\Phi_0(x) d\Phi_1(x) \\ &= E_0(x) + tE_1(x) + t^2(|d\Phi_1(x)|^2 + |d\Phi_0(x)|^2 - 2 d\Phi_0(x) d\Phi_1(x)). \end{aligned}$$

This gives

$$\int_X |d\Phi_t|^2 = e_0 + e_1 t + t^2 \int_X |d\Phi_0 - d\Phi_1|^2. \quad \square$$

We obtain the following uniqueness statement for the differential of the pluriharmonic map. While the methods do not seem to be related at all, there is a uniqueness statement for harmonic maps which has a somewhat similar form to the recent preprint [GKM08, Lemma 3.1].

THEOREM 5.6. *If  $\Phi_0$  and  $\Phi_1$  are two pluriharmonic maps of finite energy equivariant for the same representation  $\rho$ , then with appropriate determinations of charts of  $\mathcal{T}$ , we have  $d\Phi_0 = d\Phi_1$ . The continuous maps  $\Phi_t$  in the geodesic homotopy relating  $\Phi_0$  and  $\Phi_1$  are, in fact, differentiable and pluriharmonic and their differentials are all the same.*

*Proof.* The condition that  $\Phi_0$  and  $\Phi_1$  are pluriharmonic implies that  $e_2 = 0$  in the previous corollary. This is because the pluriharmonic condition means that the maps are critical points of the energy functional, so the quadratic polynomial  $e_0 + e_1 t + e_2 t^2$  has critical points at  $t = 0$  and  $t = 1$ . This implies that  $e_2 = e_1 = 0$ . In particular, with the determinations of charts as in the lemma, we have  $d\Phi_0 = d\Phi_1$ , hence  $\text{Sym}^2(\theta_0) = \text{Sym}^2(\theta_1)$ . Over the open set  $U$  then, the differentials of  $\Phi_t$  are all the same. □

We do not have a fully worked-out example, but it seems clear that the maps themselves need not be unique, rather there can be the latitude for a small interval of linear translations within  $\mathcal{T}$ . This is related to the fact that critical points of  $\theta$  need not go to vertices of  $\mathcal{T}$ .

LEMMA 5.7. *Suppose that  $f : Y \rightarrow X$  is a generically surjective map, suppose that  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, K)$  is a representation satisfying Hypothesis 4.1, and suppose that  $\Phi : \tilde{X} \rightarrow \mathcal{T}$  is a pluriharmonic map of finite energy equivariant for  $\rho$ . Then  $\Phi \circ \tilde{f}$  is a pluriharmonic map of finite energy equivariant for  $f^*(\rho)$ .*

*Proof.* The hypothesis of quasiunipotence at infinity is satisfied by  $f^*(\rho)$ , and the condition that  $f$  be generically surjective implies that  $f^*(\rho)$  is still Zariski-dense. Thus, from Proposition 4.17 we obtain the existence of an equivariant pluriharmonic map of finite energy  $\Phi' : \tilde{Z} \rightarrow \mathcal{T}$ . Choose a generically finite covering  $Z \rightarrow Y$  such that the pullbacks of the differentials  $\mathrm{Sym}^2(\theta)$  for  $\Phi$  and  $\mathrm{Sym}^2(\theta')$  for  $\Phi'$  become single-valued on  $Z$ .

We may assume that  $Z$  is smooth and has a normal crossings compactification.

Let  $\Phi_0$  and  $\Phi_1$  denote respectively the pullbacks of  $\Phi$  and  $\Phi'$  to  $Z$ . We claim that they are of finite energy. Let  $\mathrm{Sym}^2(\theta_0)$  and  $\mathrm{Sym}^2(\theta_1)$  denote the squares of their differentials. By our assumption on  $Z$  these are both squares of differential forms  $\alpha_0$  and  $\alpha_1$ , respectively, on  $Z$ . By Lemma 5.3,

$$\alpha_0^2, \alpha_1^2 \in H^0(\mathrm{Sym}^2(\Omega_{\bar{Z}}^1; \frac{1}{2} \log D_Z)).$$

This implies by a local calculation that  $\alpha_0, \alpha_1$  extend to holomorphic forms on the compactification of  $Z$ . In turn this means that  $\Phi_0$  and  $\Phi_1$  have finite energy.

Now, Theorem 5.6 says that the squares of their differentials are the same, in other words  $\mathrm{Sym}^2(\theta_0) = \mathrm{Sym}^2(\theta_1)$ . This implies that  $\mathrm{Sym}^2(\theta') = f^*\mathrm{Sym}^2(\theta)$  on  $Y$ . As  $|\mathrm{Sym}^2(\theta')|$  is integrable on  $Y$ , the same holds for  $|f^*\mathrm{Sym}^2(\theta)|$  which means that  $\Phi \circ f$  has finite energy.  $\square$

COROLLARY 5.8. *There is a finite quasiprojective ramified covering  $Z \rightarrow X$  such that over  $Z$ , the pullback of  $\Phi$  is a finite-energy pluriharmonic map, and the determination of a direction of the edges of the tree can be made and the differential becomes a holomorphic section  $\alpha \in H^0(\bar{Z}, \Omega_{\bar{Z}}^1)$ . The pluriharmonic map is constant on the leaves of the foliation defined by the real one-form  $\Re\alpha$ .*

The fact that the pluriharmonic map is not constant implies that the differential is not identically zero, so  $\mathrm{Sym}^2(\theta)$  is not identically zero, and on the covering  $Z$  the form  $\alpha$  is non-zero.

### 5.3 Geometry of the pluriharmonic map

We now discuss in a preliminary way some of the real geometry of this picture. This section is optional.

LEMMA 5.9. *In the situation of Corollary 5.8, the periods of the real one-form  $\Re\alpha$  are integers. The associated class  $[\Re\alpha] \in H^1(\bar{Z}, \mathbb{Z})$  corresponds to the map*

$$\varphi : \bar{Z} \rightarrow S^1 = \mathcal{T}'/\mathrm{SL}(2, K)',$$

where  $\mathcal{T}'$  is the subtree of  $\mathcal{T}$  consisting of edges whose interior is touched by  $\Phi$ , and we denote by  $\mathrm{SL}(2, K)' \subset \mathrm{SL}(2, K)$  the subgroup of transformations preserving  $\mathcal{T}'$  and preserving the orientations of all edges of  $\mathcal{T}'$ . This restricts to a map on the open set  $Z$  and the representation  $\rho$  on  $\pi_1(Z)$  factors through the subgroup  $\mathrm{SL}(2, K)'$ .

*Proof.* The choice of a determination of  $\alpha$  corresponds to a choice of direction for the edges of  $\mathcal{T}'$ , so  $\rho : \pi_1(Z) \rightarrow \mathrm{SL}(2, K)'$ . Now we can form the quotient metric space  $S^1 = \mathcal{T}'/\mathrm{SL}(2, K)'$  and the pluriharmonic map descends to give a map  $\varphi$  that is well defined on  $Z$ . The periods of  $\Re\alpha$  correspond to signed lengths of images of loops by  $\varphi$  so they are clearly integers. In particular, the periods of  $\Re\alpha$  on  $\bar{Z}$  are integers so the Albanese map projects to a map  $\varphi$  defined on the compactification  $\bar{Z}$ .  $\square$

In the situation of the lemma, the map  $\varphi$  factors as the composition of the Albanese map of  $\bar{Z}$  with the linear map whose differential is  $\Re\alpha$ :

$$Z \rightarrow \mathrm{Alb}(\bar{Z}) \rightarrow S^1.$$

We can also perform a ‘real Stein factorization’ of the map  $\varphi$  through a map

$$Z \rightarrow G \rightarrow S^1,$$

where  $G$  is a graph. Note in the case when  $Z$  is non-compact, the graph  $G$  may be non-Hausdorff, which is what leads to monodromy around the points at infinity.

The graph  $G$  is a version of the *Reeb graph* of a Morse function [Ree46]. This notion came to our attention in a talk given by B. Falcidieno. It has been cited recently in [CSSB04, CEHNP04, EP03].

There is a distinguished point on  $S^1$  corresponding to the image of the vertices of  $\mathcal{T}'$ . However, the vertices of  $G$  might not go to this point. In the case where no vertex of  $G$  goes to the distinguished point, then one can compose with small rotations of  $S^1$  which then lift, and obtain a family of translations of the pluriharmonic map  $\Phi$ . This is why we said above that it looked like the map was not necessarily unique. To get an example, one would have to get an example where the graph  $G$  can map to  $S^1$  sending vertices to points other than the distinguished one.

It would be good to have more information on how the representation  $\rho$  determines the covering  $Z$ , the cohomology class  $[\Re\alpha]$ , the map  $\varphi$  and its factorization through the graph  $G$ , and the lifting of  $\varphi$  to the pluriharmonic map  $\Phi$ . Ideally one would like to be able to calculate everything explicitly, whenever we are given  $\rho$  by matrices.

### 5.4 Lefschetz theory for the spectral foliation

After pulling back to the covering  $Z$  given by Corollary 5.8, we have the following situation. We have an irreducible smooth quasiprojective variety  $Z$  with smooth normal-crossings compactification  $\overline{Z}$ . The lift to  $Z$  of the multivalued spectral form  $\alpha \in H^0(\overline{Z}, \Omega_{\overline{Z}}^1)$  is a non-zero holomorphic one-form on the compactification. Let  $A$  be the minimal abelian variety from which  $\alpha$  comes (that is,  $A = \text{Alb}(\overline{Z})/B$ , where  $B$  is the maximal sub-abelian variety on which  $\alpha$  vanishes). Let  $Z'$  and  $\overline{Z}'$  be the coverings of  $Z$  and  $\overline{Z}$  defined by fiber product with the universal covering  $V \rightarrow A'$ .

Let  $\mathcal{F}$  be the foliation defined by  $\alpha$  on  $Z'$ . We have a map  $Z' \rightarrow V$ , and  $\alpha$  corresponds to a linear form on  $V$ , which we denote by  $g : V \rightarrow \mathbb{C}$ . For  $t \in \mathbb{C}$  we obtain a linear subspace  $P_t \subset V$  and  $g^{-1}(t) \subset \overline{Z}'$  is its pullback to  $\overline{Z}'$ . Let

$$D' := \overline{Z}' - Z'$$

be the covering of the complementary divisor.

LEMMA 5.10. *In the above situation, the coverings  $\overline{Z}'$  and  $Z'$  are connected, and  $\pi_1(Z')$  contains the commutator subgroup of  $\pi_1(Z)$ .*

*Proof.* The covering  $Z' \rightarrow Z$  is an infinite Galois covering whose Galois group is the quotient of  $\pi_1(Z)$  given by the surjective morphism  $\pi_1(Z) \rightarrow \pi_1(A)$ . In particular,  $Z'$  is connected and the Galois group of  $Z'/Z$  is abelian, hence  $\pi_1(Z')$  contains the commutator subgroup of  $\pi_1(Z)$ .  $\square$

The following is the basic Lefschetz theorem which we need. It is a generalization to the quasiprojective case of the statement of [Sim93] which was done for the projective case of the present project.

THEOREM 5.11. *Suppose that the complex dimension of the image of  $Z$  in  $A$  is at least two. Then if  $t \in \mathbb{C}$  is a general value (that is, not one of a countable number of singular values), the inverse images  $g^{-1}(t)$  are connected and their fundamental groups surject onto  $\pi_1(Z')$ .*

*Proof.* We first treat a simplified special case which is then easily seen to give the general case.

*Special case.* Suppose that the following additional conditions hold:  $\dim(Z) = 2$ ; the morphism from  $\overline{Z}$  onto its image  $\overline{Z}_A \subset A$  is a resolution of singularities; there is a Weil divisor  $D_A \subset \overline{Z}_A$  containing the singular set of  $\overline{Z}_A$ , such that  $D$  is the inverse image of  $D_A$  in  $\overline{Z}$ ; and the map  $Z \rightarrow \overline{Z}_A - D_A$  is

an isomorphism. In particular, we may consider this last isomorphism as an equality and view  $Z$  as an open subset of the smooth points of  $\overline{Z}_A \subset A$ .

Now as explained in [Sim93], in this situation the only zeros of  $\alpha$  on  $Z$  lie over points in  $A$ . We can also look at zeros of the restriction of  $\alpha$  to  $D$ . Since  $D$  is a curve, there are two cases: a zero of  $\alpha|_D$  is either a point or it is a union of components of  $D$ . However, if  $D_i$  is a component of  $D$  such that  $\alpha|_{D_i} = 0$ , then the albanese variety of  $D_i$  would be an abelian subvariety on which  $\alpha$  vanishes. By the construction of  $B$  and  $A$ , this means that  $D_i$  maps to a point in  $A$ . We conclude that there is a finite set of points  $p_j \in \overline{Z}_A$  containing the zeros  $\alpha|_Z$  and the images of the zeros of  $\alpha|_D$ .

We can now apply the same argument as in [Sim93]. We do not give the details of the choice of nested neighborhoods and so forth necessary to make things rigorous: this follows the same principles as in [Sim93].

Fix a collection of relatively compact neighborhoods of these points  $\overline{N}_A(p_j) \subset \overline{Z}_A$ .

Away from these neighborhoods we can choose vector fields of bounded length on  $\overline{Z}$ , preserving  $Z$  that is, tangent to  $D$ , whose images by  $\alpha$  are constant (real and imaginary fields on  $\mathbb{C}$ ).

Following these vector fields outside of the given neighborhoods, we can lift a given homotopy in  $\mathbb{C}$  in a way which takes us only within a bounded region of the non-compact  $\overline{Z}'$ . This allows us to solve the Lefschetz problem outside of the neighborhoods, see [Sim93].

Thus, we can restrict our attention to one of the neighborhoods  $\overline{N}_A = \overline{N}_A(p_j)$ . Let  $\overline{N}_1$  denote its inverse image in  $\overline{Z}$ , and let  $N_1 = \overline{N}_1 \cap Z = \overline{N}_1 - \overline{N}_1 \cap D$  be its intersection with  $Z$ . This might be disconnected because the image  $\overline{Z}_A$  is not necessarily normal. Let  $N$  be a connected component of  $N_1$  and  $\overline{N}$  the closure of  $N$  in  $\overline{N}_1$ .

The inverse image of  $\overline{N}$  in  $\overline{Z}'$  is a disjoint union of isomorphic copies of the neighborhood  $\overline{N}$  itself, and each copy denoted  $\overline{N}'$  is relatively compact in  $\overline{Z}'$  (that is, the covering  $Z'/Z$  does not induce any non-trivial coverings of the neighborhoods). Since we are in a bounded region of  $\overline{Z}'$ , only a finite number of these come into play. We just have to solve the Lefschetz problem in one of the connected neighborhoods  $N' \subset \overline{N}'$ .

Note that the map  $\overline{N}' \rightarrow \overline{N}_A(p_j)$  is a resolution of singularities of one of the irreducible components of the target. Any positive-dimensional exceptional fiber is contained in the interior of  $\overline{N}'$ , in other words does not go near the boundary; this is because the exceptional components map to the point  $p_j \in \overline{N}_A(p_j)$ .

Here the situation becomes classical. We can look at integration of  $\alpha$  as defining a holomorphic map  $g$  to a disk which is a fibration over the punctured disk. In fact this comes from a function  $g_A : \overline{N}_A(p_j) \rightarrow \Delta$  which we normalize so that  $g(p_j) = 0$ . Over the punctured disk  $\Delta^*$  the map of triples

$$(\overline{N}', D \cap \overline{N}', N')_{\Delta^*} \rightarrow \Delta^*$$

is topologically locally trivial (we arrange the boundary of the neighborhood in a good way so that this holds).

Note furthermore that  $D$  intersects the fiber  $g^{-1}(0)$  only in points lying over  $p_j$ . This is because any component of  $D$  upon which the differential  $dg = \alpha$  restricts to zero, maps to an isolated point in  $A$  (which in the present case must be  $p_j$ ). Given our hypothesis that  $D_A$  contains the singular points of  $\overline{Z}_A$ , this means in particular that the fiber  $g_A^{-1}(0)$  intersects the singular set of  $\overline{N}_A(p_j)$  only at  $p_j$ .

The Lefschetz question is now localized as follows: given  $p$  and  $q$  in one fiber  $g^{-1}(t)$  of the map  $g : N' \rightarrow \Delta$  with  $t \neq 0$ , and a path  $\gamma$  from  $p$  to  $q$  inside the total space of the neighborhood  $N'$ , is  $\gamma$  homotopic to a path which stays in the fiber?

The main tool for doing this is the observation that there is at least one component of the central fiber  $g^{-1}(0)$  which is not contained in  $D$ , and is not a multiple fiber. Indeed, the fiber  $g_A^{-1}(0)$  contains a positive-dimensional component in each irreducible component of  $\overline{N}_A(p_j)$ , by purity of



the dimension of a subset defined by one equation. As noted above, away from  $p_j$  this corresponds to a curve in our open subset  $N' \cong N$ . As  $g^{-1}(0)$  is the inverse image in  $\overline{N}'$  of  $g_A^{-1}(0) \subset \overline{N}_A(p_j)$ , this means that  $g^{-1}(0)$  contains a component which is not contained in  $D$  as claimed.

Using this fact we can treat the local Lefschetz question. Let  $\xi$  be a path starting from  $p$  and going around a non-multiple component of the central fiber. Let  $N^*$  denote the complement of  $g^{-1}(0)$  in  $N'$ . Then  $\text{Path}(N'; p, q)$  is obtained from  $\text{Path}(N^*; p, q)$  by adjoining at least the relation that  $\xi$  becomes trivial. Now the fibration

$$g : N^* \rightarrow \Delta^*$$

yields an exact sequence

$$1 \rightarrow \text{Path}(g^{-1}(t); p, q) \rightarrow \text{Path}(N^*; p, q) \rightarrow \mathbb{Z} \rightarrow 0,$$

and multiplication by  $\xi$  via the action of  $\pi_1(N^*, p)$  on  $\text{Path}(N^*; p, q)$  projects to translation by one in  $\mathbb{Z}$  (this is exactly the statement that  $\xi$  is a loop around a non-multiple component; if it had gone around a component of multiplicity  $r$  then it would have projected to translation by  $r$ ). We can now prove the Lefschetz property. The path  $\gamma$  can be moved away from the central fiber so it comes from an element of  $\text{Path}(N^*; p, q)$ ; multiplying by an appropriate power of  $\xi$  yields a path in  $\text{Path}(g^{-1}(t); p, q)$ , and since  $\xi$  is homotopically trivial in  $N'$  this means that  $\gamma$  was homotopic in  $N'$  to an element of  $\text{Path}(g^{-1}(t); p, q)$ .

Adorning the argument with the niceties of [Sim93], we are done in the special case  $\dim(Z) = 2$ .

*General case.* We now return to the general situation of the statement of the theorem and show how to obtain the conclusion, once we know the special case of dimension two which has been treated already.

Suppose that  $x', y' \in g^{-1}(t) \subset Z'$  are two points, and suppose that  $\gamma$  is a path from  $x'$  to  $y'$  in  $Z'$ . We want to show that  $\gamma$  is homotopic to a path inside  $g^{-1}(t)$ . Let  $x$  and  $y$  denote the images of  $x'$  and  $y'$  in  $Z$ . Choose a complete intersection of hyperplane sections  $\overline{Y} \subset \overline{Z}$ , such that  $x, y \in Y := \overline{Y} \cap Z$ , with  $\dim(Y) = 2$ . Choose  $\overline{Y}$  of sufficiently high multidegree and general subject to these conditions (although in the case  $\dim(Z) = 2$  we just have  $\overline{Y} = \overline{Z}$ ). In particular, the image of  $Y$  in  $A$  has dimension two because by hypothesis the image of  $Z$  has dimension at least two. The map  $\pi_1(\overline{Y}) \rightarrow \pi_1(A)$  is still surjective so the covering  $Y' := Y \times_Z Z'$  is connected as in Lemma 5.10. Also, the fundamental group of  $Y$  surjects onto that of  $Z$ , so  $\pi_1(Y') \rightarrow \pi_1(Z')$  is surjective. Now  $x', y' \in Y'$ , and  $\gamma$  is homotopic to a path in  $Y'$ . Thus, it suffices to prove that  $g^{-1}(t) \cap Y'$  is connected and has fundamental group surjecting to  $\pi_1(Y')$ , for then we will be able to make  $\gamma$  homotopic to a path from  $x'$  to  $y'$  in  $g^{-1}(t) \cap Y'$ .

For the proof for  $Y'$ , we can replace  $Y$  by a smaller open subset without loss of generality, even if  $x'$  or  $y'$  is no longer in this smaller open subset. This is because connectedness and surjectivity of the map on fundamental groups are properties which are stable under passing between Zariski-open subsets and supersets. Finally, we can also change the compactification  $\overline{Y}$  without changing the statement. Now by going to a small enough open subset  $Y$  and choosing a normal crossings compactification  $\overline{Y}$ , we can obtain the special case which was treated in the first part of the proof. This completes the proof in the general case. □

We now apply the above result to the spectral data for our pluriharmonic map. Look first over  $Z$ . The differential of the pluriharmonic map is  $\pm\alpha$  for a well-defined one-form  $\alpha$ . The restriction of the harmonic map to any of the leaves  $g^{-1}(t)$  of the resulting foliation, is constant. Therefore, if  $\gamma \in \pi_1(Z)$  is a path which has a representative whose lift to the universal covering stays within a leaf, then the action of  $\gamma$  fixes the image point of the leaf.

LEMMA 5.12. *The dimension of the image of  $Z$  in  $A$  is one.*

*Proof.* Suppose to the contrary that the dimension of the image of  $Z$  in  $A$  is at least two. Then by the Theorem 5.11, for a generic value of  $t$  we have the fiber  $g^{-1}(t) \subset Z'$  which is connected and has fundamental group surjecting to  $\pi_1(Z')$ . Consider the pullback

$$\tilde{Z}_t := g^{-1}(t) \times_{Z'} \tilde{Z}.$$

It is a union of possibly infinitely many connected components denoted by  $\tilde{Z}_{t,k}$ .

Let  $\Phi : \tilde{Z} \rightarrow \mathcal{T}$  be the pluriharmonic map. The differential of  $\Phi$  is constant along the components  $\tilde{Z}_{t,k}$  by construction of  $g$  as the integral of  $\alpha = \pm d\Phi$ . Choose a component  $\tilde{Z}_{t,k}$ , which therefore maps to a single point  $q \in \mathcal{T}$  under  $\Phi$ . If  $x \in \tilde{Z}_{t,k}$  is a point, and if  $\gamma \in \pi_1(Z')$ , then the translate is again in the same component:

$$\gamma \cdot x \in \tilde{Z}_{t,k}.$$

In particular,  $\Phi(\gamma \cdot x) = q = \Phi(x)$ . The equivariance property of  $\Phi$  tells us that  $\Phi(\gamma \cdot x) = \rho(\gamma)\Phi(x)$ , so we obtain  $\rho(\gamma)q = q$ . In other words, under the assumption that the dimension of the image of  $Z$  in  $A$  is at least two, applying Theorem 5.11 we conclude that  $\pi_1(Z')$  has a fixed point in  $\mathcal{T}$  under the action  $\rho$ .

By Lemma 5.10, the commutator subgroup the commutator subgroup  $\Upsilon \subset \pi_1(Z)$  is contained in  $\pi_1(Z')$ . Thus, the assumption that the dimension of the image of  $Z$  in  $A$  is at least two would imply that the commutator subgroup  $\Upsilon$  acts with a fixed point on  $\mathcal{T}$ . By Lemma 4.18 this would mean that  $\Upsilon$  maps to the center of  $\text{SL}(2, K)$ . However, that would imply that the representation into  $\text{PSL}(2, K)$  factors through the abelianization  $\pi_1(X, x)^{\text{ab}}$  and, hence, to a commutative subgroup scheme, contradicting Zariski-denseness of  $\rho$ .

We conclude that the dimension of the image  $Z_A$  cannot be two or more. Since  $\alpha$  is non-trivial, we have  $\dim(Z_A) = 1$ . □

Putting all of the above discussion together we obtain the following theorem, see [Bea92, Cam94, CS83, Del07, Eys04, Gro89, GS92, JZ00, Kat94, Kat97, Kli03, Kol93, NR08, Siu80, Zuo94].

**THEOREM 5.13.** *Suppose that  $X$  is a quasiprojective variety, that  $K$  is a complete local field with finite residue field, that  $\rho : \pi_1(X, x) \rightarrow \text{SL}(2, K)$  is a representation whose monodromy transformations at infinity are quasiunipotent, that the image of  $\rho$  is Zariski-dense, and that the image is not contained in any compact subgroup. Then  $\rho$  factors through a DM-curve.*

*Proof.* Considering the projected representation  $\tilde{\rho}$  into  $\text{PSL}(2, K)$ , we are in the situation of Hypothesis 4.1. Apply all of the above sequence of results. We obtain a covering  $Z \rightarrow X$  and a pluriharmonic map on  $X$  which pulls back to one on  $Z$ . Let

$$Z \rightarrow Y \rightarrow A$$

be the Zariski–Stein factorization of the partial Albanese map  $Z \rightarrow A$  discussed above. By Lemma 5.12,  $Y$  is a curve. It is normal because  $Z$  is, so it is a smooth curve. The fibers of  $Z \rightarrow Y$  are contained in the leaves of the foliation defined by  $\Re\alpha$ , and the pluriharmonic map is constant on these leaves, by Corollary 5.8. Thus, the pluriharmonic map is constant on the fibers of  $Z \rightarrow Y$ .

The image of the fundamental group of the fiber is a normal subgroup  $\Upsilon \subset \pi_1(Z, z)$ . This acts on  $\mathcal{T}$  with a non-empty set of fixed points, indeed any image of a fiber by the pluriharmonic map is a point fixed by  $\Upsilon$ . So applying Lemma 4.18, we have that  $\Upsilon$  maps to the identity element of  $\text{PSL}(2, K)$ . Therefore,  $\rho_Z$  projectively factors through the map  $Z \rightarrow Y$ . By Lemma 3.1,  $\rho_Z$  factors through a DM-curve. Then by Lemma 3.6 we obtain a factorization over  $X$ . □

In the remainder of the paper, we consider a representation  $\rho : \pi_1(X, x) \rightarrow \text{SL}(2, \mathbb{C})$  which does not factor through a DM-curve. The above theorem (applied to various fields  $K$ ) will allow us to

prove that  $\rho$  is rigid and integral, and this will eventually give rise to a variation of Hodge structure corresponding to a factorization through a Shimura modular stack.

### 6. Rigidity

Suppose that  $X$  is a smooth quasiprojective variety. Fix a smooth compactification  $X \subset \overline{X}$  such that the complementary divisor  $D$  has smooth irreducible components  $D_i$  meeting at normal crossing points. For each component  $D_i$  of the divisor at infinity, let  $\gamma_i$  denote a loop going out to a point near  $D_i$  then once around  $D_i$  and back to  $x$  along the same path.

We start with some general notation. Suppose that  $G$  is a reductive group. Suppose that we have chosen closed  $\text{ad}(G)$ -invariant subsets  $C_i \subset G$  indexed by the components  $D_i$  of the divisor at infinity. The *representation space*

$$R(X, x, G, \{C_i\})$$

is defined as the closed subset of  $\text{Hom}(\pi_1(X, x), G)$  of representations  $\rho$  such that  $\rho(\gamma_i) \in C_i$ . The condition that  $C_i$  be conjugation-invariant means that this condition is well-defined independently of the choices of  $\gamma_i$ .

We can then define the *moduli stack* and *coarse moduli space* denoted, respectively, by

$$\mathcal{M}(X, G, \{C_i\}) \quad \text{and} \quad M(X, G, \{C_i\}).$$

These are the quotients of  $R(X, x, G, \{C_i\})$  by the conjugation action of  $G$ , first in the sense of algebraic stacks and then in the sense of a universal categorical quotient of affine schemes. These are the ‘Betti’ moduli spaces [LM85], there should also be ‘de Rham’ and ‘Dolbeault’ versions [BBN01, Bis97, Biq96, BY96, IIS06, Kon93, Nak96, Nit93, Sim95, SW01, Tha02, Yok93].

The main case we consider is when  $C_i$  are the closures of conjugacy classes in  $G$ . Suppose that  $\rho : \pi_1(X, x) \rightarrow G$  is a representation. Let  $C_i(\rho)$  be the closure of the conjugacy class of the image  $\rho(\gamma_i)$ . We say that a Zariski-dense representation  $\rho$  is *rigid* if it represents an isolated point in the moduli space  $M(X, G, \{C_i(\rho)\})$ , and *non-rigid* otherwise.

#### 6.1 Conjugacy classes in $\text{SL}(2)$

Specialize now to the case  $G = \text{SL}(2)$ . In this case there are basically two types of conjugacy classes: the unipotent matrices and the semisimple matrices with distinct eigenvalues. One has to add in the semisimple matrices with only one eigenvalue, which in this case means just  $\pm 1$ . The notion of ‘unipotent’ should also be extended to include  $-1$  times a unipotent matrix which we sometimes call ‘projectively unipotent’.

LEMMA 6.1. *Suppose that  $C_i$  is the closure of a conjugacy class in  $\text{SL}(2, K)$  with  $K$  an algebraically closed field of characteristic different from two. Then there are five cases:*

- (i)  $C_i = \{1\}$ ;
- (ii)  $C_i = \{-1\}$ ;
- (iii)  $C_i$  is the closure of the set of matrices conjugate to  $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ ;
- (iv)  $C_i$  is the closure of the set of matrices conjugate to  $\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$ ; or
- (v) there is  $a \neq \pm 1$  in  $K$  such that  $C_i$  is the set of matrices conjugate to  $\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$ .

Note that the conjugacy classes (i) and (ii) are contained in the closures of the conjugacy classes (iii) and (iv), whereas the conjugacy class in (v) is closed. The equations for cases (iii), (iv) and (v) are, respectively,

$$\text{Tr}(A) = 2, \quad \text{Tr}(A) = -2, \quad \text{Tr}(A) = a + a^{-1}.$$

The equations for cases (i) and (ii) are not obtained by traces but rather directly by the equations  $A = 1$  or  $A = -1$ . In any case one sees explicitly the equations to be imposed on  $R(X, x, \text{SL}(2))$  to cut out  $R(X, x, \text{SL}(2), \{C_i\})$ .

A closed subset  $C_i$  is said to be *quasiunipotent* if there is  $n$  such that  $A^n$  is unipotent for any  $A \in C_i$ . The closure of a conjugacy class is quasiunipotent if and only if the conjugacy class is that of a quasiunipotent matrix. A conjugacy-class closure which is quasiunipotent is contained in one of the cases (i), (ii), (iii), (iv), or (v) with  $a$  a root of unity. Thus, if the  $C_i$  are quasiunipotent, then the equations defining  $R(X, x, \text{SL}(2), \{C_i\})$  are defined over any ring extension of  $\mathbb{Z}$  which contains the roots of unity in question.

Special to the case of  $\text{SL}(2)$  is the fact that a quasiunipotent conjugacy class consists either of matrices which are either of finite order (case (v)) or projectively unipotent, that is, a central multiple of a unipotent matrix (cases (iii) and (iv)).

### 6.2 Zariski density

We need to keep track of the Zariski-density hypothesis as we reduce and localize. First recall the connected subgroups of  $\text{SL}(2, k)$ .

LEMMA 6.2. *Suppose that  $k$  is an algebraically closed field. The connected algebraic subgroups of  $\text{SL}(2, k)$  are the following:*

- the trivial subgroup;
- a torus, conjugate to the subgroup of diagonal matrices;
- a unipotent subgroup conjugate to the group of strictly upper-triangular matrices;
- a solvable subgroup conjugate to the group of upper-triangular matrices; or
- the whole group.

Furthermore, if  $H \subset \text{SL}(2, k)$  is a strict algebraic subgroup which is not finite, then either  $H/H^o$  is of order at most two or  $H^o$  is conjugate to the unipotent subgroup,  $H$  is contained in the associated solvable subgroup, and  $H/H^o$  is a finite cyclic subgroup of the torus.

*Proof.* The first part comes from the Levi decomposition and the fact that the only reductive groups smaller than  $\text{SL}(2)$  are tori. For the last statement, note that  $H$  normalizes its connected component  $H^o$ . If  $H^o$  is the diagonal torus, its normalizer is the semidirect product of  $H^o$  with the group of transpositions which has order two. If  $H^o$  is the solvable (but not unipotent or toric) subgroup, then it is equal to its own normalizer. The remaining case is where  $H^o$  is the unipotent subgroup; its normalizer is the solvable subgroup and  $H/H^o$  is contained in the torus (quotient of the solvable subgroup by the unipotent subgroup). □

COROLLARY 6.3. *Still assuming  $k = \bar{k}$ , suppose that  $H \subset \text{SL}(2, k)$  is a strict algebraic subgroup containing an element whose eigenvalues are elements of infinite order in  $k^*$ . Then  $H$  is either a torus, the solvable subgroup, or the normalizer of a torus.*

*Proof.* The condition on eigenvalues means that  $H$  must contain a torus. Thus,  $H^o$  is either a torus or the solvable subgroup. In the first case,  $H$  is either the torus or its normalizer (which contains the torus with index two). In the second case,  $H = H^o$  since the solvable group is its own normalizer. □

We consider a subgroup  $H$  containing an element whose eigenvalues are of infinite order. We would like a good algebraic way of distinguishing between the cases  $H = \text{SL}(2, k)$ , and the other three possibilities mentioned in the corollary. For this, note that for the connected subgroups mentioned in Lemma 6.2, the traces of elements detect only the diagonal part and this is commutative. Thus, for example,  $\text{Tr}(ABAB) = \text{Tr}(A^2B^2)$  for these subgroups whereas this is not generically true. The other

possible subgroup, the normalizer of a torus, is an extension of order two, which we can get around by looking at the squares of matrices. This leads to the following Katz-style criterion [Kat90], also somewhat similar to [CS83, Corollary 1.2.2]. We thank M. Larsen for comments about it.

LEMMA 6.4. *Suppose that  $k$  is a field and  $\Gamma \subset \text{SL}(2, k)$  is a subgroup. Suppose that  $\Gamma$  contains an element whose eigenvalues are of infinite order in  $k^*$ . Then  $\Gamma$  is Zariski-dense if and only if there exist elements  $\alpha, \beta \in \Gamma$  such that*

$$\text{Tr}(\alpha^2\beta^2\alpha^2\beta^2) - \text{Tr}(\alpha^4\beta^4) \neq 0.$$

*Proof.* We reduce to the case  $k = \bar{k}$  so that there is no problem with twisted forms of tori. If  $\Gamma \subset \text{SL}(2, k)$  is not Zariski-dense in  $\text{SL}(2, \bar{k})$  then the Zariski closure of  $\Gamma$  over  $\bar{k}$  is a subgroup  $G \subset \text{SL}(2, \bar{k})$  preserved by  $\text{Gal}(\bar{k}/k)$ , in particular it comes from a subgroup over  $k$ . Thus,  $\Gamma$  is not Zariski-dense over  $k$ . The other direction is easy, so Zariski-density is preserved by going from  $k$  to  $\bar{k}$ . The criterion given in the statement of the lemma does not depend on  $k$  so it is preserved too. Thus, we may assume  $k = \bar{k}$ .

Now we get to the proof. If  $H \subset \text{SL}(2, k)$  is a torus, the normalizer of a torus, or a solvable subgroup, and if  $\alpha, \beta \in H$  then

$$\text{Tr}(\alpha^2\beta^2\alpha^2\beta^2) - \text{Tr}(\alpha^4\beta^4) = 0.$$

In view of Lemma 6.2, this shows that if  $\alpha, \beta \in \Gamma$  exist as in the statement, then  $\Gamma$  is Zariski-dense.

Suppose now that  $\Gamma$  is Zariski-dense. It is easy to see that  $\Gamma \times \Gamma$  is Zariski-dense in  $\text{SL}(2, k) \times \text{SL}(2, k)$ . On the other hand, the set of pairs of matrices  $(A, B) \in \text{SL}(2, k) \times \text{SL}(2, k)$  where the identity  $\text{Tr}(A^2B^2A^2B^2) = \text{Tr}(A^4B^4)$  holds is a Zariski-closed subset. In order to prove that  $\Gamma \times \Gamma$  goes outside of this subset it suffices to exhibit any pair of matrices  $(A, B) \in \text{SL}(2) \times \text{SL}(2)$  which is not in the subset. Furthermore, we can do this after going to the algebraic closure.

The subset of matrices in  $\text{SL}(2, \bar{k})$  which are squares is Zariski-dense, so again its product is Zariski-dense in the product. Thus, it suffices to exhibit a pair of matrices  $(A, B) \in \text{SL}(2, k) \times \text{SL}(2, k)$  such that  $\text{Tr}(ABAB) \neq \text{Tr}(A^2B^2)$ . For this, set

$$A := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$B := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

with  $\text{Tr}(ABAB) = 7 \neq 6 = \text{Tr}(A^2B^2)$  in any characteristic. □

LEMMA 6.5. *Suppose that  $\rho$  is a Zariski-dense representation. Then  $\rho$  is rigid in the above sense, that is, it represents an isolated point in the moduli space  $M(X, \text{SL}(2), \{\bar{C}_i(\rho)\})$ , if and only if there is no non-isotrivial family of representations all having the same conjugacy classes at infinity, going through  $\rho$ .*

*Proof.* We have a map

$$R(X, \text{SL}(2), \{\bar{C}_i(\rho)\}) \rightarrow M(X, \text{SL}(2), \{\bar{C}_i(\rho)\}).$$

As is well known, the points of the moduli space  $M$  represent  $S$ -equivalence classes of representations in  $R$ . In particular, if  $\rho$  is an irreducible representation as is the case here, any representation  $S$ -equivalent to  $\rho$  is actually isomorphic to  $\rho$ . Thus, the fiber of the map  $R \rightarrow M$  over the point  $[\rho]$  consists only of conjugates of  $\rho$ , in other words the fiber is set-theoretically an orbit of  $\text{SL}(2)$ . The point  $[\rho] \in M$  is an isolated point if and only if the fiber  $\text{SL}(2) \cdot \rho$  is a connected component of  $R$ .

If this is the case, and if  $\{\rho_t\}$  is a continuous family of representations with  $\rho_0 = \rho$  and all having the same conjugacy classes at infinity, then we obtain a path in  $R(X, \text{SL}(2), \{\bar{C}_i(\rho)\})$  going

through  $\rho$ . Since the orbit  $SL(2) \cdot \rho$  is a connected component, this means that all of the  $\rho_t$  are contained in the orbit so the family is isotrivial. This shows one direction of the lemma.

In the other direction, suppose that  $[\rho]$  is not an isolated point of  $M$ . Then the orbit  $SL(2) \cdot \rho$  is not a connected component of  $R$ . However, it is a closed subset. Therefore, there is a path  $\{\rho_t\}$  in  $R$  which starts at  $\rho_0$  in the orbit, and goes out of the orbit. Up to changing the basis of the underlying space we may assume  $\rho_0 = \rho$ . The family is non-isotrivial. We just have to see that the conjugacy classes at infinity stay the same. The monodromy transformations at infinity  $\rho_t(\gamma_i)$  are families in  $\overline{C}_i(\rho)$ . However, by definition  $\overline{C}_i(\rho)$  is the closure of the conjugacy class  $C_i(\rho)$  of  $\rho(\gamma_i)$ , in particular  $C_i(\rho)$  (which is a locally closed subset of the group) is an open subset of  $\overline{C}_i(\rho)$ . On the other hand,  $\rho_0(\gamma_i) = \rho(\gamma_i)$  is in this open subset. So, for  $t$  close to 0, we still have  $\rho_t(\gamma_i) \in C_i(\rho)$  which is the condition we needed.  $\square$

LEMMA 6.6. *If  $V$  is a rigid local system and  $\sigma : \mathbb{C} \rightarrow \mathbb{C}$  is an automorphism of  $\mathbb{C}$ , then  $V^\sigma$  is also rigid.*

The proof is standard.

### 6.3 Nonfactorization implies rigidity

To motivate the proof of the next main theorem, we note a corollary of Lemma 2.8.

COROLLARY 6.7. *The subset of points in the space of  $SL(2, \mathbb{C})$  representations of  $\pi_1(X, x)$  corresponding to representations which do not factor through a hyperbolic DM curve is open.*

*Proof.* By Lemma 2.8 there are only a finite number of maps from  $X$  to hyperbolic DM-curves. For each of these maps, the set of representations which factor is given by a closed subset.  $\square$

THEOREM 6.8. *Suppose that  $\rho : \pi_1(X, x) \rightarrow SL(2, \mathbb{C})$  is a Zariski-dense representation with quasi-unipotent monodromy at infinity. If  $\rho$  does not factor through a map from  $X$  to a DM-curve, then it is rigid in the above sense.*

*Proof.* Suppose that we have a representation  $\rho$  which does not factor through a DM-curve, and which is not rigid. We derive a contradiction. Note that  $\rho$  does not projectively factor through an orbicurve (Lemma 3.1).

We first give the short if somewhat heuristic version of the argument. From Corollary 6.7 there is an open subscheme of the moduli scheme corresponding to representations which do not factor through DM-curves. We can take an open subset to obtain an irreducible positive-dimensional open subscheme whose representations do not factor. This open set may be considered as having finite type over  $\mathbb{Z}$ . Then there is a curve defined over a finite field mapping into here. The trace functions of the monodromy elements are not all constant, in particular there is a point on the curve where at least one trace function has a pole. Taking the completion at this point we obtain a representation

$$\hat{\rho} : \pi_1(X, x) \rightarrow SL(2, \mathbb{F}_q((t)))$$

whose image is not contained in a compact subgroup. Using the criterion of Lemma 6.4 we can assume that  $\hat{\rho}$  is Zariski-dense. This is an elementary example of the more general problem of independence of  $\ell$  for monodromy groups such as considered in [LP92]. By Theorem 5.13 the representation  $\hat{\rho}$  would have to factor through a DM curve, contradicting our choice of open set.

To get a more accurate version of this proof, we fill in explicitly some of the equations for the properties preserved when going to the completion at a finite point. The argument uses what is known as a ‘spread’ [GG03].

Start by enumerating the possible factorization maps through hyperbolic orbicurves. Denote these by  $f_i : X \rightarrow Y_i$  for a finite set of indices  $i \in I$ . Let  $\varphi_i : \pi_1(X, x) \rightarrow \Gamma_i$  be the corresponding

maps on fundamental groups, and let  $K_i$  be the kernel of  $\varphi_i$ . The fact that  $\rho$  does not projectively factor through  $f_i$  translates by saying that  $\rho(K_i)$  is not contained in the center of  $\mathrm{SL}(2, \mathbb{C})$ .

Note that  $K_i$  is a normal subgroup of  $\pi_1(X, x)$  so the Zariski closure of  $\rho(K_i)$  is a normal subgroup of the full image of  $\rho$  which is, by assumption,  $\mathrm{SL}(2, \mathbb{C})$ . The only normal subgroups are the identity, the center, and the group itself. Therefore,  $\rho|_{K_i}$  is Zariski-dense in  $\mathrm{SL}(2, \mathbb{C})$ . In particular there exists an element  $k_i \in K_i$  such that

$$\mathrm{Tr}(\rho(k_i)) \neq \pm 2.$$

The non-factorization hypothesis translates into the existence of such a  $k_i$  for each  $i \in I$ .

The hypothesis that  $\rho$  is non-rigid means that there is a curve of representations passing through  $\rho$  and projecting to a non-trivial curve in the moduli space of representations. Explicitly write down this curve as follows. We may assume it is affine, and even (applying Noether normalization after going to an open set if necessary) that it is finite over  $\mathbb{A}^1$  with a coordinate ring of the form

$$A = \frac{\mathbb{C}[y]}{(y^m + a_{m-1}y^{m-1} + \dots + a_0)}.$$

Our path of representations is a representation

$$\rho_A : \pi_1(X, x) \rightarrow \mathrm{SL}(2, A)$$

and there is a point  $p : A \rightarrow \mathbb{C}$  such that  $\rho_A \otimes_{A,p} \mathbb{C} = \rho$ .

Recall that Procesi’s theorem [Pro76] says that the traces of the images of group elements give an embedding of the moduli space, see also [CS83]. Therefore, the hypothesis that our curve is non-trivial in the moduli space means that there is an element  $\gamma \in \pi_1(X, x)$  such that  $\mathrm{Tr}(\rho_A(\gamma))$  is not contained in the ground field  $\mathbb{C}$ . Concretely, the elements of  $A$  are written as sums  $u_{m-1}y^{m-1} + \dots + u_0$  so the non-constancy of this element means that one of its coordinates  $u_j$  is non-zero for  $0 < j \leq m - 1$ . Write this condition as  $u_j(\mathrm{Tr}(\rho_A(\gamma))) \neq 0$ .

Similarly, the non-factorization hypothesis states that  $\mathrm{Tr}(\rho_A(k_i)) \neq \pm 2$  for all  $i \in I$ .

We can specialize our curve of representations so that it is defined over a finite extension field  $L$  of  $\mathbb{Q}$  (in other words we specialize the coefficients  $a_i$  of the equations for  $A$  as well as the coefficients of the matrix coefficients of  $\rho_A$ ). We can do this and still keep the non-factorization conditions that  $\mathrm{Tr}(\rho_A(k_i)) \neq \pm 2$  and the non-rigidity condition that  $u_j(\mathrm{Tr}(\rho_A(\gamma))) \neq 0$ .

Now our curve of representations can also be defined over a finite extension ring  $R$  of  $\mathbb{Z}$  (in general  $R$  will be the localization of the ring of integers  $\mathcal{O}_L$  at a finite number of primes). Then we can find a prime ideal  $\mathfrak{p} \subset R$  such that the non-factorization and non-rigidity conditions hold modulo  $\mathfrak{p}$ . We obtain a curve of representations  $\rho_B : \pi_1(X, x) \rightarrow \mathrm{SL}(2, B)$  such that

$$B = \frac{\mathbb{F}_q[y]}{(y^m + b_{m-1}y^{m-1} + \dots + b_0)},$$

such that  $\mathrm{Tr}(\rho_B(k_i)) \neq \pm 2$  and  $u_j(\mathrm{Tr}(\rho_B(\gamma))) \neq 0$ . We may assume that the characteristic of the finite field  $\mathbb{F}_q$  is different from two.

Note that  $\mathrm{Spec}(B)$  is a curve over the finite field  $\mathbb{F}_q$  and we may view the element  $\mathrm{Tr}(\rho_B(\gamma))$  as a non-constant function on this curve. Complete at a point at infinity on this curve, choosing a point with the property that the function  $\mathrm{Tr}(\rho_B(\gamma))$  has a pole. The completed field, which may be seen as the completion of  $B$  with respect to the non-archimedean norm given by the order of pole at the point, will be denoted by  $\widehat{B}$ . It is a complete local field of the type envisioned in Hypothesis 4.1. The ring of integers  $\mathcal{O}_{\widehat{B}}$  is a discrete valuation ring with finite residue field  $\mathbb{F}_q$ , so

$$\mathcal{O}_{\widehat{B}} \cong \mathbb{F}_q[[t]], \quad \widehat{B} \cong \mathbb{F}_q((t)).$$

We have a representation

$$\rho_{\widehat{B}} : \pi_1(X, x) \rightarrow \mathrm{SL}(2, \widehat{B}) \cong \mathrm{SL}(2, \mathbb{F}_q((t))).$$

The fact that  $\mathrm{Tr}(\rho_{\widehat{B}}(k_i)) \neq \pm 2$  means that  $\rho_{\widehat{B}}$  does not projectively factor through the orbicurve  $f_i : X \rightarrow Y_i$ . This holds for all  $i \in I$ , so  $\rho_{\widehat{B}}$  does not factor through an orbicurve. On the other hand, the fact that  $\mathrm{Tr}(\rho_B(\gamma))$  has a pole at the point where we took the completion means that  $\mathrm{Tr}(\rho_{\widehat{B}}(\gamma))$  is not in  $\mathcal{O}_{\widehat{B}}$ .

We have  $\rho_{\widehat{B}}(\gamma) \in \mathrm{SL}(2, \widehat{B})$  a matrix with eigenvalues  $\beta^{\pm 1}$  for  $\beta \in \overline{\widehat{B}}$ , therefore  $\rho_{\widehat{B}}(\gamma^n)$  has eigenvalues  $\beta^{\pm n}$ . In particular, if  $\mathrm{Tr}(\rho_{\widehat{B}}(\gamma))$  has a pole of order at least one, then  $\mathrm{Tr}(\rho_{\widehat{B}}(\gamma^n))$  has a pole of order at least  $n$ . This shows that  $\rho_{\widehat{B}}$  does not have image in any compact subgroup of  $\mathrm{SL}(2, \widehat{B})$ .

This situation leads to a contradiction with Theorem 5.13. Before getting there, we still have to verify the hypothesis that  $\rho_{\widehat{B}}$  have Zariski-dense image. This is done much as above, using the explicit criterion of Lemma 6.4. Note that the image of  $\rho$  was not finite, so it contained an element of infinite order. By Lemma 6.4, there exist  $\alpha, \beta \in \pi_1(X, x)$  such that

$$\mathrm{Tr}(\rho(\alpha^2 \beta^2 \alpha^2 \beta^2)) - \mathrm{Tr}(\rho(\alpha^4 \beta^4)) \neq 0.$$

In our process of reducing modulo a prime, we may choose a prime which does not divide this non-zero quantity. Then generalizing along a curve and completing we still obtain a representation  $\rho_{\widehat{B}}$  with the property that

$$\mathrm{Tr}(\rho_{\widehat{B}}(\alpha^2 \beta^2 \alpha^2 \beta^2)) - \mathrm{Tr}(\rho_{\widehat{B}}(\alpha^4 \beta^4)) \neq 0.$$

On the other hand, our construction of  $\rho_{\widehat{B}}$  was such that its image contains an element of infinite order. Thus, Lemma 6.4 again applies (choose our prime to have characteristic different from two). We conclude that the image of  $\rho_{\widehat{B}}$  is Zariski-dense.

We can now apply Theorem 5.13 to obtain a contradiction, completing the proof of the theorem. □

*Remark 3.* The above proof constitutes a new proof of the factorization result of [Sim91] in the projective case. Delzant seems to have found a similar proof, as he has indicated to us in correspondence. Many of the details—fundamentally elementary—going into the proof, stem from wanting to refer only to trees of finite type. The theory of Gromov and Schoen should in principle allow us to prove the result more directly for a complex curve of representations going to infinity, but this would involve considerations of non-locally compact trees which we have preferred to avoid.

Alternatively, one could adapt to the quasiprojective (and quasiunipotent monodromy) case the archimedean argument of [Sim91]. This type of thing is currently being done by Mochizuki in a much wider context [Moc02, Moc07]. For example, he obtains the theorem that any representation can be deformed to a complex variation of Hodge structure. We felt that in the context of the present paper, it would be more interesting to deduce Theorem 6.8 from the Gromov–Schoen theory of harmonic maps to trees, since this provides a unified approach to the two basic results (rigidity as above, and integrality in the next section). We cannot entirely avoid referring to the archimedean case: in §8 we state the theorem that rigid representations are complex variations of Hodge structure, which requires the theory of harmonic maps to symmetric spaces.

### 7. Integrality

Suppose that  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, \mathbb{C})$  is a representation. Following Bass [Bas80] we say that  $\rho$  is *integral* if for every  $\gamma \in \pi_1(X, x)$  the trace  $\mathrm{Tr}(\rho(\gamma))$  is an algebraic integer. Bass proved the following fundamental characterization.



LEMMA 7.1. A Zariski-dense representation  $\rho$  has integral traces if and only if there exists a number field  $L$  (that is, a finite extension of  $\mathbb{Q}$ ) with its ring of algebraic integers  $\mathcal{O}_L \subset L$ , an embedding  $\eta : L \rightarrow \mathbb{C}$ , a rank-two projective  $\mathcal{O}_L$ -module  $V_{\mathcal{O}_L}$ , and a representation

$$\rho_{\mathcal{O}_L} : \pi_1(X, x) \rightarrow \mathrm{SL}(V_{\mathcal{O}_L})$$

such that

$$\rho_{\mathcal{O}_L} \otimes_{\mathcal{O}_L, \eta} \mathbb{C} \cong \rho.$$

See [Bas80], for the proof.

LEMMA 7.2. Suppose that  $V$  is a local system, and suppose that  $Y \rightarrow X$  is a finite étale covering such that  $V|_Y$  is integral. Then  $V$  is integral.

*Proof.* A local system is integral if and only if for any element  $\gamma \in \pi_1(X, x)$ , all of the eigenvalues of  $\rho_V(\gamma)$  are algebraic integers. Under the hypotheses of the lemma, if  $\gamma \in \pi_1(X, x)$ , then there is an  $n$  such that  $\gamma^n$  comes from  $\pi_1(Y, y)$  (where  $y$  is a lift of the basepoint to  $Y$ ). Then the eigenvalues of  $\rho_V(\gamma^n)$  (which are the  $n$ th powers of the eigenvalues of  $\rho_V(\gamma)$ ) are algebraic integers. This implies that the eigenvalues of  $\rho_V(\gamma)$  are algebraic integers (indeed an equation of the form  $z^n = \alpha$  is an integral equation for  $z$ , if  $\alpha$  is an algebraic integer).  $\square$

The main result about integrality is the following. It is pretty well known, see the work of Jost and Zuo [JZ00], Eyssidieux [Eys04], Katzarkov [Kat94, Kat97], Klingler [Kli03], Napier and Ramachandran [NR08] and others which are based on the theory of Gromov–Schoen [GS92].

THEOREM 7.3. If  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, \mathbb{C})$  is a Zariski-dense representation with quasiunipotent monodromy at infinity, and  $\rho$  does not factor through a DM-curve, then  $\rho$  is integral.

*Proof.* Suppose that  $\rho$  does not factor. By Theorem 6.8,  $\rho$  is rigid. This implies that it can be defined over an algebraic number field  $L$  (which we may assume is finite over  $\mathbb{Q}$ ). Assume that it is not integral. Then the trace of some monodromy element is not an algebraic integer, and in particular it is not integral at some prime of  $L$ . Taking the completion at this prime, we get a Zariski-dense representation into  $\mathrm{SL}(2, K)$  where  $K$  is a complete local field, and the representation does not go into a compact subgroup because the trace of some element is not an integer of  $K$ . The residue field of  $K$  is finite because the integers of  $L$  are of finite type over  $\mathbb{Z}$ . The monodromy transformations at infinity are the same as for the original  $\rho$ , so they are quasiunipotent. Therefore, Theorem 5.13 applies: the representation into  $\mathrm{SL}(2, K)$  factors through a DM-curve. By invariance of factorization under change of base field, Corollary 3.8, the original representation factors through a DM-curve contradicting our hypothesis. Therefore,  $\rho$  is integral.  $\square$

### 7.1 Hypergeometric cases

There is not a perfect dichotomy between the properties of factorization through DM-curves and rigidity. There are some cases where these overlap, corresponding to rigid local systems on DM-curves. For Zariski-dense representations on orbicurves, rigidity can be determined by a dimension count which depends only on the local monodromy data. Katz’s algorithm [Kat96], in principle, gives a way to determine which rigid monodromy data can come from representations. Furthermore, when the data do come from representations, his algorithm provides an explicit construction of the representation as motivic. As a corollary, rigid representations on DM-curves are automatically integral.

In the case of local systems of rank two, the only rigid cases correspond to hypergeometric equations; their motivic expression is basically the classical integral expression of hypergeometric functions [Del86, Kat90, Kat96]. In particular, the result that rigidity implies integrality can be understood explicitly in this case.

Suppose that  $Y$  is a smooth compact curve with points  $y_1, \dots, y_k$  and basepoint  $y$  distinct from the marked points. Suppose that  $\rho : \pi_1(Y - \{y_i\}, y) \rightarrow \mathrm{SL}(r, \mathbb{C})$  is a Zariski-dense representation, which is rigid among those with fixed conjugacy classes at the punctures.

LEMMA 7.4. *In the above situation,  $Y \cong \mathbb{P}^1$ .*

*Proof.* A dimension count shows that the only potentially problematic case which needs to be discussed is that of representations on an elliptic curve, with prescribed monodromy around a single puncture. This yields the equation

$$aba^{-1}b^{-1} \in C$$

for a fixed conjugacy class  $C$ . The space of solutions is divided by the conjugation action of  $\mathrm{SL}(2)$ . The dimension of the space of solutions, divided by the conjugation action, is therefore at least equal to

$$2 \dim(\mathrm{SL}(2)) - \dim(\mathrm{SL}(2)) - (\mathrm{codim}_{\mathrm{SL}(2)} C) = \dim(C).$$

In particular, as soon as  $C$  is positive-dimensional, the representation cannot be rigid. The only zero-dimensional conjugacy classes are  $\{1\}$  and  $\{-1\}$ . In the first case the image is abelian; and in the second case it is an extension of an abelian group by  $\{\pm 1\}$ . Neither of these can be Zariski-dense subgroups of  $\mathrm{SL}(2)$ . Therefore, this case cannot arise, so  $Y$  has to be rational.  $\square$

Now we proceed with a dimension count. Note that the naive dimension count actually gives the dimension of the moduli space at a Zariski-dense point, in this case [Kat96]. Let  $C_1, \dots, C_k$  denote the conjugacy classes of the representation at the points  $y_i$ . Here we are including orbifold points because the space of representations on the orbifold is the same as the space of representations on the Zariski-open subset with prescribed conjugacy classes (in the case of orbifold points, the conjugacy classes will be of finite order).

Recall that the space of representations is the space of solutions of  $a_1 \cdots a_k = 1$  with  $a_i \in C_i$ , the whole up to a global conjugation (and on Zariski-dense points the conjugation action is faithful). The virtual dimension of the moduli space is therefore

$$v := \dim(C_1) + \cdots + \dim(C_k) - 2 \dim(\mathrm{SL}(2)).$$

Rigidity happens for Zariski-dense representations exactly when  $v = 0$ .

Recall that the dimensions of the conjugacy classes for the cases given by Lemma 6.1 are as follows: for (i) and (ii),  $\dim(C) = 0$ ; whereas for (iii), (iv) and (v),  $\dim(C) = 2$ . Together with the above we conclude the following statement.

LEMMA 7.5. *If  $\rho$  is a rigid rank-two representation on a DM-curve, then there are exactly three orbifold or singular points with conjugacy classes different from  $\{1\}$  or  $\{-1\}$ .*

The points with conjugacy classes  $\{-1\}$  can be chosen freely; they can also combine with the three main points (resulting in an interchange between (iii) and (iv), or taking the negative of the eigenvalue in case (v)). These operations correspond to tensoring with rank-one local systems on  $Y$  (in fact, the Prym local systems of hypergeometric curves over  $Y$ ). From now on we ignore these shifts, and use the term ‘unipotent’ for cases (iii) and (iv).

If all three conjugacy classes are of type (v), then we are in the case of the classical hypergeometric equation. It is well known that for eigenvalues which are roots of unity, these hypergeometric local systems exist, and are motivic (hence, integral).

We give explicit calculations for the cases where one or more of the conjugacy classes is unipotent, showing that the monodromy representations are integral. Note by [Kat96] that if a rigid representation exists, it is uniquely determined by the monodromy data; thus it suffices to write down a

single integral representation for each given collection of monodromy data. In a similar way, if there exists a reducible representation with given conjugacy classes, then no irreducible representations will exist. In this case we must multiply one of the matrices by  $-1$ . Also for these calculations, note that a matrix with trace  $\pm 2$  which is not  $\pm$  the identity, must be in conjugacy classes (iii) or (iv) (that is, non-trivially unipotent).

For three unipotent matrices, one or three must have eigenvalues  $-1$  otherwise an irreducible solution will not exist. In this case we have the solution

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix} = \begin{pmatrix} -3 & 1 \\ -4 & 1 \end{pmatrix}.$$

For two unipotent matrices and one matrix of type (v) with eigenvalue  $\alpha$ , note that

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} = \begin{pmatrix} 1+b & 1 \\ b & 1 \end{pmatrix}.$$

If  $b = \alpha + \alpha^{-1} - 2$ , then the trace on the right will be  $\alpha + \alpha^{-1}$  so we obtain a solution. For any root of unity  $\alpha \neq \pm 1$  the coefficient  $b$  is an algebraic integer so this gives an integral solution. Note also that there is a solution for  $-\alpha$ , so by tensoring with a rank-one system we can also treat unipotent conjugacy classes of type (iv).

Finally, for the case of one unipotent matrix and two matrices of type (v) with eigenvalues  $\alpha$  and  $\beta$ , look at

$$\begin{pmatrix} \alpha & 1 \\ 0 & \alpha^{-1} \end{pmatrix} \begin{pmatrix} \beta & 0 \\ x & \beta^{-1} \end{pmatrix} = \begin{pmatrix} \alpha\beta + x & \beta^{-1} \\ x\alpha^{-1} & (\alpha\beta)^{-1} \end{pmatrix}.$$

Up to tensoring by a rank-one Prym system (corresponding to changing the sign of one of  $\alpha$  or  $\beta$ ) we can assume that we want the trace on the right to be equal to two. Note that in any case the matrix on the right is not the identity so if its trace is two it will be of type (iii). This gives the equation

$$x = 2 - (\alpha\beta + (\alpha\beta)^{-1}).$$

If  $\alpha$  and  $\beta$  are roots of unity and  $\alpha\beta \neq 1$  then this has an integral solution. If  $\alpha\beta = 1$ , then there is a non-irreducible representation so there can be no irreducible solution (however, if one of the matrices is multiplied by  $-1$  then there is a solution).

We have concluded the proof of the following result.

**LEMMA 7.6.** *If  $\rho$  is a Zariski-dense rank-two rigid representation on a DM-curve, then it is integral.*

*Remark 4.* Our main result, Theorem 11.2 below, implies that in all of the above cases, the representation is motivic. Thus, we are spared the trouble of writing down an explicit motivic presentation for the cases of one or more unipotent conjugacy classes. Such presentations certainly exist in the literature, see [DR00, Kat90, Kat96]. Darmon discusses motivic properties of rigid local systems on  $\mathbb{P}^1$  in an arithmetic setting [Dar00].

### 8. Variations of Hodge structure

We recall here the classical definitions concerning complex, real, and integral variations of Hodge structure (VHSs). In this section, let  $V$  be a complex local system over  $X$ . We think of it as a  $C^\infty$  vector bundle with flat connexion  $\nabla$ . A structure of  $\mathbb{C}$ VHS of weight  $w$  on  $V$  is the data of a decomposition of  $C^\infty$  vector bundles  $V = \bigoplus_{p+q=w} V^{p,q}$  satisfying the Griffiths identities

$$\nabla(V^{p,q}) \subset A^{1,0}(X; V^{p,q}) \oplus A^{1,0}(X; V^{p-1,q+1}) \oplus A^{0,1}(X; V^{p,q}) \oplus A^{0,1}(X; V^{p+1,q-1}).$$

A real VHS arises when  $V$  has a real structure  $V_{\mathbb{R}}$  (that is, a real local system whose tensor product with  $\mathbb{C}$  is  $V$ ), such that  $V^{p,q} = \overline{V^{q,p}}$ . Finally an integral VHS is a real VHS together with an integral local system  $V_{\mathbb{Z}} \subset V_{\mathbb{R}}$  whose extension by scalars to  $\mathbb{R}$  is  $V_{\mathbb{R}}$ .

If  $V$  is a CVHS of weight  $w$ , a *hermitian polarization* of  $V$  is a sesquilinear  $\nabla$ -invariant form  $\Phi(\cdot, \cdot)$  on  $V$ , such that  $\Phi$  is  $(-1)^1$ -sesquisymmetric

$$\Phi(u, v) = (-1)^w \overline{\Phi(v, u)}$$

the different Hodge subspaces are orthogonal

$$\Phi(u, v) = 0 \quad \text{for } u \in V^{p,q}, v \in V^{r,s}, (p, q) \neq (r, s);$$

and such that the positivity condition

$$(\sqrt{-1})^{p-q} \Phi(u, u) > 0, \quad u \in V^{p,q}$$

holds.

A *real polarization* on a real variation  $(V, V_{\mathbb{R}})$  is a hermitian polarization  $\Phi$  such that  $\Phi(u, v) \in \mathbb{R}$  whenever  $u, v \in V_{\mathbb{R}}$ . An *integral polarization* on a integral variation  $(V, V_{\mathbb{Z}})$  is an hermitian polarization  $\Phi$  such that  $\Phi(u, v) \in \mathbb{Z}$  whenever  $u, v \in V_{\mathbb{Z}}$ .

In the weight-one case which interests us here, our conventions for the polarization say that  $\Phi$  should be a hermitian antisymmetric form, in particular  $\sqrt{-1}\Phi$  is an indefinite unitary form with  $\sqrt{-1}\Phi(u, u)$  positive for  $u \in V^{1,0}$  and negative for  $u \in V^{0,1}$ . If  $\Phi$  is a real or integral polarization, then it induces an alternating form on  $V_{\mathbb{R}}$  or  $V_{\mathbb{Z}}$ , in particular  $\Phi(u, u) = 0$  for  $u \in V_{\mathbb{R}}$ . The hermitian antisymmetric form  $\Phi$  on the complexification is uniquely determined by  $\Phi|_{V_{\mathbb{R}} \otimes V_{\mathbb{R}}}$  and the positivity condition holds on the Hodge subspaces which do not intersect  $V_{\mathbb{R}}$ .

Recall [Del71, Del79] that, if we make the convention that  $V^{p,q}$  is to be non-zero only when  $p, q \geq 0$ , then an integral polarizable weight-one VHS is the same thing as an algebraic smooth family of abelian varieties over  $X$ .

**THEOREM 8.1.** *Suppose that  $V$  is an irreducible rank-two local system with quasiunipotent monodromy at infinity. If  $V$  is rigid as defined in § 6, then it underlies a complex VHS of weight one.*

*Proof.* This is [Car92, Corollary 1], see also Jost and Zuo [JZ96], then Mochizuki [Moc07, §§ 21–25]. Recall that the proof goes as follows, starting with the analogue of the discussion in §§ 4–5. There is a finite energy harmonic metric  $h$  on  $V$  with the property that flat sections on sectors going towards points at infinity, have sub-polynomial growth (and the same for the dual bundle). This metric is unique up to multiplication by a positive scalar. The flat connection  $\nabla$  decomposes as

$$\nabla = \partial + \bar{\partial} + \theta + \bar{\theta}$$

and the metric is pluriharmonic, that is, the Bochner formulae [Cor88] still hold [Car92, JZ96, Moc07]. For example,  $\theta$  is a holomorphic endomorphism for the Higgs complex structure  $\bar{\partial}$ .

This fits into a more general picture as follows. Recall from [Moc02, Moc07, § 22.1] that a flat bundle with pluriharmonic metric  $h$  is *tame* if the eigenvalues of  $\theta$  extend as bounded multivalued holomorphic sections of the bundle  $\Omega^1_X(\log D)$  of logarithmic differential forms. In this case the *residues* of the eigenvalues are a well-defined collection of numbers called the KMS-spectrum, and  $h$  is *purely imaginary* if the residues are purely imaginary. A tame harmonic bundle is purely imaginary if and only if flat sections on sectors going towards points at infinity have sub-polynomial growth. Any irreducible local system admits a unique (up to positive scalar) tame pluriharmonic metric such that the residues of  $\theta$  along  $D$  are purely imaginary [Moc07, § 25.28]. The imaginary parts of the eigenvalues of the residue correspond to the logarithms of the norms of the eigenvalues of the monodromy.

In our case, the quasiunipotence condition on the monodromy of  $\nabla$  means that the eigenvalues of the residue of  $\theta$  are zero, in other words the eigenvalues of  $\theta$  have poles of order strictly less than one. An equivalent formulation is that the harmonic map has finite energy with respect to our Poincaré-type metric on the base [JZ96].

For a tame pluriharmonic metric satisfying the conditions of the previous paragraph, we obtain *norm estimates* analogous to those for variations of Hodge structure. In particular, when measured with respect to a metric restricted from  $\overline{X}$ ,

$$|\theta|_h \leq C \sum_i |\log|z_i||^k$$

on a neighborhood of a point of  $D$  where the components  $D_i$  are defined by coordinates  $z_i$ . The norm estimates in the higher-dimensional case are one of the main conclusions of [Moc07], although for our purposes it would suffice to restrict to a curve transverse to  $D$  and refer to [Sim90].

Consider a sequence of small loops  $\gamma^\epsilon$  with diameter  $\epsilon \rightarrow 0$  based at points  $x^\epsilon$  approaching some smooth point of the divisor  $D$  (these loops can be in a transverse curve). Compare the holonomy transformations of the unitary part

$$\text{hol}(\partial + \overline{\partial}; \gamma^\epsilon) \in \text{U}(V_{x^\epsilon}) \subset \text{GL}(V_{x^\epsilon})$$

with the monodromy transformations which are the (not necessarily unitary) holonomy transformations of  $\nabla$ . The norm estimates plugged into an integral estimate for the holonomy imply that

$$|\text{hol}(\nabla; \gamma^\epsilon) - \text{hol}(\partial + \overline{\partial}; \gamma^\epsilon)| \leq C\epsilon |\log \epsilon|^k.$$

It follows that the characteristic polynomials of  $\text{hol}(\partial + \overline{\partial}; \gamma^\epsilon)$  have a limit as  $\epsilon \rightarrow 0$  and this limit is the characteristic polynomial of the monodromy transformation  $\text{hol}(\nabla; \gamma^\epsilon)$  around the smooth point of  $D$  in question.

We have now put in place all of the ingredients necessary for the same proof as in the compact case. Bochner pluriharmonicity implies that for  $t \in \text{U}(1) \subset \mathbb{C}^*$  the connection

$$\nabla_t := \partial + \overline{\partial} + t\theta + \overline{t}\overline{\theta}$$

is again flat. This is a continuous family of connections, and the unitary pieces of  $\nabla_t$  with respect to  $h$  are all the same  $\partial + \overline{\partial}$ . So the paragraph on holonomy above yields that the characteristic polynomials of the monodromy transformations of  $\nabla_t$  are the same as for  $\nabla$  (in particular, the monodromy is still quasiunipotent).

Therefore, we can apply the rigidity hypothesis (cf. Lemma 6.5), and conclude that the local systems corresponding to  $(V, \nabla_t)$  are all isomorphic. This means that there are automorphisms  $\alpha_t : V \rightarrow V$  with  $\nabla_t = \alpha_t \circ \nabla \circ \alpha_t^{-1}$ . The fixed metric  $h$  is pluriharmonic for all of the connections  $\nabla_t$ . It is also tame, and indeed the  $\nabla_t$ -flat sections again have subpolynomial growth, as can be seen using the above norm estimates for the difference  $|\nabla_t - \nabla| = |(t - 1)\theta + (\overline{t} - 1)\overline{\theta}|$  plus an integral estimate for the solutions of  $\nabla_t(v) = 0$ .

The eigenvalues of  $t\theta$  have poles of order strictly less than one, and the harmonic map given by  $h$  with the flat connection  $\nabla_t$  has finite energy in the Poincaré-type metric, as its differential is  $t\theta$  which will have the same  $L^2$  norm as  $\theta$ .

Unicity of the pluriharmonic metric on the flat bundle  $(V, \nabla_t)$ —with respect to any one of the properties of subpolynomial growth, zero residual eigenvalues for  $\theta$ , or finite energy—means that  $\alpha_t$  preserves  $h$  up to a scalar constant, and modifying  $\alpha_t$  we can assume it preserves  $h$ . Thus,

$$\alpha_t \circ \theta \circ \alpha_t^{-1} = t\theta,$$

and decomposing  $V$  into generalized eigenspaces for  $\alpha_t$  yields the required decomposition to give a structure of VHS just as in the compact case [Hit95, Sim92]. The above intertwining formula says that  $\theta$  goes from the  $u$ -generalized eigenspace of  $\alpha_t$ , to the  $tu$ -generalized eigenspace.

In our rank-two case, there are at most two generalized eigenvalues of  $\alpha_t$ ; if there are exactly two which differ by multiple of  $t$  we obtain a Hodge decomposition with two adjacent Hodge types, otherwise the term  $\theta$  linking them is zero so the connection is unitary. If there is only one generalized

eigenvalue then the connection is unitary too. In the case of two adjacent Hodge types we can assume that the Hodge numbers are  $h^{1,0} = h^{0,1} = 1$ , whereas in the unitary case we can choose  $h^{1,0} = 2$  or  $h^{1,0} = 2$ . The harmonic metric  $h$  yields an hermitian polarization  $\Phi$  by the formula

$$\Phi(v, v) = \sqrt{-1}^{q-p} h(v, v), \quad v \in V^{p,q}. \quad \square$$

**COROLLARY 8.2.** *Suppose that  $V$  is a rank-two local system with quasiunipotent monodromy at infinity, whose monodromy representation is Zariski-dense. If  $V$  does not factor through a DM-curve, then for any automorphism  $\sigma$  of  $\mathbb{C}$ , the conjugate local system  $V^\sigma$  underlies a complex VHS of weight one.*

*Proof.* By Theorem 6.8,  $V$  is rigid. Therefore, any conjugate is also rigid, and by Theorem 8.1 the conjugates all underly complex VHSs.  $\square$

When a representation is integral and all of the Galois conjugates come from complex VHSs, then adding them together we obtain a  $\mathbb{Z}$ -VHS.

**LEMMA 8.3.** *If  $V_{\mathcal{O}_L}$  is a local system of projective  $\mathcal{O}_L$ -modules for an algebraic number field  $L$ , such that for every embedding  $\eta : L \rightarrow \mathbb{C}$  the resulting complex local system  $V^\eta$  underlies a complex VHS, then  $V_{\mathcal{O}_L}^{\oplus 2}$  considered as a local system of  $\mathbb{Z}$ -modules underlies a  $\mathbb{Z}$ -VHS. If each factor can be associated to a variation of weight one, then the  $\mathbb{Z}$ -VHS can be assumed to have weight one and, in particular, it comes from a family of abelian varieties.*

See [Sim92] for a proof.

As a corollary we get a first statement which approaches our classification statement.

**COROLLARY 8.4.** *Suppose that  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, \mathbb{C})$  is a representation with quasiunipotent monodromy at infinity, which does not factor through a map to a DM-curve. Then there is a family of abelian varieties over  $X$  such that  $\rho$  is a direct factor of the underlying complex monodromy representation.*

*Proof.* By Corollary 8.2 we are in the situation of Lemma 8.3, so  $\rho$  is a complex direct factor in the monodromy of a weight-one  $\mathbb{Z}$ -VHS; this is the monodromy representation of the corresponding family of abelian varieties.  $\square$

The task in the remainder of the paper is to analyse more carefully the kind of variation of Hodge structure which can be made to occur here, and show that it corresponds to a factorization of the representation through a Shimura modular stack.

## 9. Polydisk Shimura modular stacks

Moduli for families of abelian varieties is obtained through the theory of *Shimura varieties*. In general, a Shimura variety will have a tautological representation of  $\pi_1$  into some algebraic group. Since we are interested in rank-two representations, we consider the special case where this algebraic group is some form of  $\mathrm{SL}(2)$ . The classical case which has been the most extensively studied is that of *Hilbert modular varieties*, corresponding to  $\mathrm{SL}(2, F)$  where  $F$  is a totally real field [Dar00, Fre90, Gor02, HU74, McM03, Rap78, Van88].

It turns out that the Hilbert modular case is very slightly overly restrictive for our purposes. For a rigid and integral representation thanks to a lemma explained to us by M. Larsen, we can assume that the representation goes into  $\mathrm{SL}(2, L)$  where  $L$  is a totally imaginary quadratic extension of a totally real field  $F$ , see § 10. However, the universal example which we consider in this section shows that, in general, the traces will not necessarily lie in  $F$ , so it is not always possible to reduce

to  $SL(2, F)$ . We thus fall into the more general realm of the theory of Shimura varieties, more specifically unitary Shimura varieties [Lar92, Pap00, RZ96] which have recently come to be known as ‘Shimura varieties of PEL type’ [Wed99]. Our case is midway between the Hilbert modular case and the general unitary case, because the fact that the group is  $SL(2)$  means that the universal covering will be a product of disks. For this reason, and because of the apparent lack of any standard terminology, we propose to call these ‘polydisk Shimura varieties’. As with the case of DM-curves, it is more convenient, when trying to obtain good factorization statements, to ignore level structure and consider moduli stacks.

Thus, we review a version of Shimura’s construction which we call *polydisk Shimura modular DM-stacks*. Aside from the restriction to the group  $SL(2)$ , this is standard material [Del71, Del79, Kot92, LP92, Pap00, RZ96, Wed99], but it seems like a good idea to have a discussion which corresponds to our specific situation.

Let  $L$  be a totally imaginary extension of a totally real field  $F$ . Let  $\mathcal{O}_L$  be the ring of algebraic integers in  $L$ . Let  $P$  be a projective  $\mathcal{O}_L$ -module of rank two, and put  $P_L := P \otimes_{\mathcal{O}_L} L$ .

Suppose that  $\sigma : L \rightarrow \mathbb{C}$  is an embedding. Let  $P^\sigma := P_L \otimes_L \mathbb{C}$  be the tensor product using  $\sigma$  as structural map.

Let  $\iota : L \rightarrow L$  be the complex conjugation map. It is independent of the embedding (indeed it is the generator for the order-two group  $\text{Gal}(L/F)$ ), so for any embedding  $\sigma$  we have

$$\sigma(\iota x) = \overline{\sigma x}.$$

Let  $\overline{\sigma} = \sigma \circ \iota$  denote the embedding composed with complex conjugation.

Now we assume that we are given an  $\iota$ -hermitian antisymmetric form  $\Phi$  on  $P_L$ . This means

$$\Phi : P_L \times P_L \rightarrow L,$$

such that  $\Phi(u, v) = -\iota\Phi(v, u)$ , with  $\Phi$  separately additive in each variable and such that  $\Phi(au, v) = a\Phi(u, v)$  thus also  $\Phi(u, av) = (\iota a)\Phi(u, v)$ . The collection  $(L, P, \Phi)$  is basically what is known as a *PEL datum* [Wed99].

Let  $\mathcal{U}(P_L, \Phi)$  denote the group of  $L$ -linear transformations of  $P_L$  which preserve  $\Phi$ , and let  $\mathcal{U}(P, \Phi)$  denote the subgroup of those transformations which also preserve  $P \subset P_L$ . Let  $\mathcal{SU}(P_L, \Phi)$  and  $\mathcal{SU}(P, \Phi)$  denote the subgroups of transformations of determinant one.

If  $\sigma : L \rightarrow \mathbb{C}$  is an embedding then we obtain a complex hermitian antisymmetric form on  $P^\sigma := P_L \otimes_L \mathbb{C}$  defined by the condition that

$$\Phi^\sigma(u^\sigma, v^\sigma) = \sigma\Phi(u, v)$$

where  $u^\sigma$  and  $v^\sigma$  denote the images of  $u$  and  $v$  in  $P^\sigma$ .

A *Hodge structure on  $(P_L, \Phi)$*  is the data  $H$  for each embedding  $\sigma$  of a Hodge decomposition

$$P^\sigma = H_\sigma^{1,0} \oplus H_\sigma^{0,1},$$

such that this decomposition is polarized by the form  $\Phi^\sigma$  (recall that this means that  $\Phi^\sigma$  is positive imaginary on  $H_\sigma^{1,0}$  and negative imaginary on  $H_\sigma^{0,1}$ , and that these two spaces are orthogonal), and such that the decompositions on  $P^\sigma$  and  $P^{\overline{\sigma}}$  are complex conjugate. Use the following notation for this complex conjugation condition. We have an antilinear isomorphism

$$\overline{(\cdot)} : P^\sigma \leftrightarrow P^{\overline{\sigma}}, \quad \overline{u^\sigma} := u^{\overline{\sigma}}$$

(recall that  $\overline{\sigma} := \sigma\iota$ ), and using this operation we require that

$$H_\sigma^{p,q} = \overline{H_\sigma^{q,p}}.$$

To see why such structures exist, we need the following lemma.

LEMMA 9.1. *Suppose that*

$$P^\sigma = \bigoplus H_\sigma^{p,q}$$

*is a Hodge decomposition polarized by  $\Phi^\sigma$ . Then setting*

$$H_\sigma^{p,q} := \overline{H_\sigma^{q,p}}$$

*we obtain a Hodge decomposition*

$$P^{\bar{\sigma}} = \bigoplus H_{\bar{\sigma}}^{p,q}$$

*for  $P^{\bar{\sigma}}$  polarized by  $\Phi^{\bar{\sigma}}$ .*

*Proof.* Suppose that  $u, v \in P$  so that  $u^\sigma, v^\sigma \in P^\sigma$ . Then

$$\Phi^{\bar{\sigma}}(\overline{u^\sigma}, \overline{v^\sigma}) = \bar{\sigma}\Phi(u, v) = \overline{\sigma\Phi(u, v)} = \overline{\Phi^\sigma(u^\sigma, v^\sigma)}.$$

This shows that, in general, for  $x, y \in P^\sigma$  we have

$$\Phi^{\bar{\sigma}}(\overline{x}, \overline{y}) = \overline{\Phi^\sigma(x, y)}.$$

Now if  $x \in H_\sigma^{p,q}$ , then  $\overline{x} \in H_{\bar{\sigma}}^{q,p}$  and to obtain the polarization condition for  $H_{\bar{\sigma}}^{q,p}$  we would like to show that

$$(\sqrt{-1})^{q-p}\Phi^{\bar{\sigma}}(\overline{x}, \overline{x}) > 0.$$

By the previous formula, the quantity on the left is equal to

$$(\sqrt{-1})^{q-p}\overline{\Phi^\sigma(x, x)} = \overline{(\sqrt{-1})^{p-q}\Phi^\sigma(x, x)}.$$

The polarization condition for  $H_\sigma^{p,q}$  says that

$$(\sqrt{-1})^{p-q}\Phi^\sigma(x, x) > 0$$

which yields the desired polarization condition for  $H_{\bar{\sigma}}^{q,p}$ . The condition about orthogonality is proved similarly. □

Our forms  $\Phi^\sigma$  are hermitian antisymmetric, giving hermitian symmetric forms denoted by  $\sqrt{-1}\Phi^\sigma$ .

Say that  $\sigma$  is *positive*, *mixed*, or *negative*, respectively, if  $\sqrt{-1}\Phi^\sigma$  is a positive-definite, indefinite, or negative-definite unitary form. Note that  $\bar{\sigma}$  is, respectively, negative, mixed, or positive when  $\sigma$  is positive, mixed, or negative (a special case of the above lemma).

If  $\sigma$  is positive (respectively negative) then we are forced to set  $H_\sigma^{1,0} := P^\sigma$  (respectively,  $H_\sigma^{0,1} := P^\sigma$ ) so in these cases the Hodge decompositions are uniquely determined. In the mixed case, each Hodge subspace has dimension one and the space of such decompositions for a given  $\sigma$  is the unit disk. The complex conjugacy condition determines the Hodge structure on  $P^{\sigma^c} = P^{\bar{\sigma}}$  once it is given for  $P^\sigma$ , and the complex conjugate Hodge structure on  $P^{\sigma^c}$  is polarized by  $\Phi^{\sigma^c}$  if and only if the original one was polarized by  $\Phi^\sigma$ , by Lemma 9.1. Therefore, the space of Hodge decompositions of  $(P_L, \Phi)$  is isomorphic to a product of disks, with the number of factors equal to half the number of mixed embeddings for  $\Phi$ .

If  $X$  is a smooth analytic variety, then a *VHS of type  $(P, \Phi)$  over  $X$*  is a local system  $V$  of  $\mathcal{O}_L$  modules whose stalks are isomorphic to  $P$ , together with an antihermitian form  $\Phi_V$  on  $V$  such that for any  $x \in X$  the stalk  $(V_x, \Phi_{V,x})$  is isomorphic to  $(P, \Phi)$  (and, in particular,  $V$  is a local system of projective rank-two  $\mathcal{O}_L$ -modules), together with a family of Hodge decompositions of the  $(V_x, \Phi_{V,x})$  in the above sense which for each  $V^\sigma$  forms a complex variation of Hodge structure polarized by  $\Phi_V^\sigma$ . This notion can be extended in an obvious way to the case where the base  $X$  is a smooth analytic DM-stack.



Let  $\mathcal{D}(P, \Phi)$  denote the period domain parametrizing Hodge decompositions on  $(P, \Phi)$  as above. The group  $\mathcal{U}(P, \Phi)$  acts on  $\mathcal{D}(V_L, \Phi)$ . Put

$$\mathfrak{H}^{\text{an}}(P, \Phi) := \mathcal{D}(P, \Phi)/\mathcal{U}(P, \Phi).$$

**THEOREM 9.2.** *The analytic stack  $\mathfrak{H}^{\text{an}}(P, \Phi)$  is a smooth analytic DM-stack (that is, an orbifold). This means, in particular, that the stabilizers are finite. It admits a tautological VHS of type  $(P, \Phi)$  denoted by  $\mathcal{V}$ . This represents the functor of VHSs of type  $(P, \Phi)$  in the following sense: if  $X$  is any smooth analytic variety (or analytic DM-stack) and  $(V, \Phi_V)$  is a VHS of type  $(P, \Phi)$  over  $X$ , then there is a map  $f : X \rightarrow \mathfrak{H}^{\text{an}}(P, \Phi)$  and an isomorphism*

$$\alpha : (V, \Phi_V) \cong f^*(\mathcal{V}, \Phi_{\mathcal{V}}).$$

Furthermore, given two maps  $f, g : X \rightarrow \mathfrak{H}^{\text{an}}(P, \Phi)$ , the set of isomorphisms between  $f$  and  $g$  maps isomorphically to the set of isomorphisms between  $f^*(\mathcal{V}, \Phi_{\mathcal{V}})$  and  $g^*(\mathcal{V}, \Phi_{\mathcal{V}})$ .

**THEOREM 9.3.** *There exists an algebraic DM-stack  $\mathfrak{H}(P, \Phi)$  and an isomorphism between its associated analytic space and the complex DM-stack  $\mathfrak{H}^{\text{an}}(P, \Phi)$  defined above. The algebraic stack represents the functor of VHSs of type  $(P, \Phi)$  over algebraic varieties (with the same statement as before, so we do not repeat it here). In particular, if  $X$  is any smooth algebraic DM-stack, then any analytic map  $X^{\text{an}} \rightarrow \mathfrak{H}^{\text{an}}(P, \Phi)$  comes from an algebraic map (and the analytic and algebraic isomorphisms between algebraic maps are the same).*

This is a classical result of Baily–Borel type. In the Hilbert modular case see Rapoport [Rap78, Theorem 1.20]. Rapoport attributes the materiel of his § 1 to Deligne. For Shimura varieties of PEL type, Wedhorn [Wed99] refers to Kottwitz [Kot92] who states a moduli problem in terms of abelian varieties with appropriate structure, in our case that would mean purely imaginary multiplication by  $L$ .

The consequence about unicity of the algebraic structure seems even now a bit surprising. It basically comes from the distance-decreasing property for variations of Hodge structure, which implies that any analytic map is distance decreasing for the Poincaré metric.

The stacks  $\mathfrak{H}(P, \Phi)$  appearing in Theorem 9.3 are called *polydisk Shimura modular DM-stacks*. Note that there is a tautological representation

$$\pi_1(\mathfrak{H}(P, \Phi)) \rightarrow \mathcal{U}(P, \Phi) \hookrightarrow \text{SL}(2, L)$$

which gives one tautological representation  $\pi_1(\mathfrak{H}(P, \Phi)) \twoheadrightarrow \text{SL}(2, \mathbb{C})$  for each embedding  $\sigma : L \hookrightarrow \mathbb{C}$ .

A necessary remark is that our terminology is meant to suggest that these are the standard examples of Shimura varieties whose universal covering is a polydisk. However, we do not claim that *every* Shimura variety of PEL type whose universal covering is a polydisk is of the above form (or otherwise Hilbert modular). It would be good to clarify this.

Ideally one should also discuss the notion of *level structure* here, of a type generically denoted by  $N$  for the purposes of the present paragraph. This allows one to get rid of the stack structure by looking at finite étale coverings  $\mathfrak{H}(P, \Phi, N) \rightarrow \mathfrak{H}(P, \Phi)$  by covering stacks which are actually quasiprojective varieties. One would then need to define the notion of *VHS of type  $(P, \Phi)$  with level structure* to correspond to morphisms into the polydisk Shimura modular varieties with level structure  $\mathfrak{H}(P, \Phi, N)$ . For a given map  $X \rightarrow \mathfrak{H}(P, \Phi)$  one could try to describe the explicit finite étale covering of  $X$  which would lift to a map into  $\mathfrak{H}(P, \Phi, N)$ . Such a discussion would have the benefit of giving a factorization result in the world of varieties rather than in the world of stacks. However, choices would have to be made and the factorization would only take place over a covering of  $X$ . Just as we have amply seen in the case of curves, the formulation in terms of stacks gives a much more natural and cleaner factorization statement.

10. Improvements for a rigid integral local system

Let  $\Upsilon := \pi_1(X, x)$ . Suppose  $\rho : \Upsilon \rightarrow \mathrm{SL}(2, \mathbb{C})$  is a rigid, integral Zariski-dense representation. As was pointed out by M. Larsen (see [Sim92, Lemma 4.8]), we can assume after simultaneous conjugation that  $\rho$  is defined over a totally imaginary quadratic extension  $L$  of a totally real field  $F$ . In other words  $\rho$  comes from  $\rho_L : \Upsilon \rightarrow \mathrm{SL}(2, L)$  by extension of scalars via an embedding  $\eta : L \rightarrow \mathbb{C}$ .

Furthermore, by Bass–Serre theory [Bas80] the fact that  $\rho$  is integral means that there is a rank-two projective  $\mathcal{O}_L$ -module  $P$  with  $P \otimes_{\mathcal{O}_L} L \cong L^2$  such that  $\rho_L$  comes from a representation  $\rho_P : \Upsilon \rightarrow \mathrm{SL}(P)$ .

We adopt the same notation as in the previous section, for example  $\iota$  denotes the complex conjugation on  $L$ . If  $\sigma : L \rightarrow \mathbb{C}$  is any embedding then we obtain a complex vector space  $P^\sigma := P \otimes_{\mathcal{O}_L} \mathbb{C}$  and a representation  $\rho^\sigma : \Upsilon \rightarrow \mathrm{SL}(P^\sigma)$  as in the previous section. For the given embedding  $\sigma = \eta$  this gives back the original representation. If  $\rho$  was rigid then all of the  $\rho^\sigma$  are rigid too. Thus, they all correspond to complex variations of Hodge structure.

This is almost the data we need in order to define a Shimura modular VHS and obtain a map from  $X$  into a polydisk Shimura modular stack  $\mathfrak{H}(P, \Phi)$ . However, we have to discuss the question of the polarization.

10.1 Uniqueness of the polarization

Let  $V$  be a vector space of dimension two over  $L$ , with an action of  $\Upsilon$ . Define the *conjugate representation*  $V^\iota$  as follows. The underlying additive group is the same as that of  $V$ , but the scalar multiplication operation  $\mu$  is defined by

$$\mu_{V^\iota}(u, x) := \mu_V(\iota u, x).$$

The action of  $\Upsilon$  is the same as on  $V$ , and it is again by  $L$ -linear endomorphisms.

A sesquilinear form  $\Phi$  on  $V$  is a map  $\Phi : V \times V \rightarrow L$  which is separately additive in each variable and which satisfies

$$\Phi(ax, y) = a\Phi(x, y),$$

and

$$\Phi(x, ay) = (\iota a)\Phi(x, y).$$

It is *symmetric* (respectively *antisymmetric*) if

$$\Phi(y, x) = \iota\Phi(x, y) \quad (\text{respectively } \Phi(y, x) = -\iota\Phi(x, y)).$$

We say that  $\Phi$  is  $\Upsilon$ -invariant if  $\Phi(\gamma x, \gamma y) = \Phi(x, y)$  for any  $\gamma \in \Upsilon$  and  $x, y \in V$ .

Define the *tensor corresponding to*  $\Phi$  denoted by

$$\mathbf{t}\Phi \in V^* \otimes_L (V^\iota)^*,$$

to be the unique element such that

$$\mathbf{t}\Phi \cdot x \otimes y = \Phi(x, y)$$

for any  $x \in V$  and  $y \in \iota V$  (the latter being the same as saying  $y \in V$ ); and where the product is the natural contraction

$$(V^* \otimes_L (V^\iota)^*) \times (V \otimes_L (V^\iota)) \rightarrow L.$$

There is an action of  $\Upsilon$  on  $V^* \otimes_L (V^\iota)^*$  and  $\Phi$  is  $\Upsilon$ -invariant if and only if  $\mathbf{t}\Phi$  is fixed by this action.

Define the  $\iota$ -antilinear involution  $C : V^* \otimes_L (V^\iota)^* \rightarrow V^* \otimes_L (V^\iota)^*$  with  $C^2 = 1$  by

$$C(e \otimes f) = -f \otimes e.$$

If  $\mathbf{t}\Phi = e \otimes f$  this means that  $\Phi(x, y) = e(x)f(y)$ . Define  $\Phi^C$  by  $\mathbf{t}(\Phi^C) = C(\mathbf{t}\Phi)$ . Then

$$\mathbf{t}\Phi^C = -f \otimes e$$

so  $\Phi^C(x, y) = -f(x)e(y) = -\Phi(y, x)$ . By linearity this extends to all  $\Phi$  so

$$\Phi^C(x, y) = -\Phi(y, x).$$

In particular,  $\Phi^C = \Phi$  or, equivalently,  $\mathbf{t}\Phi$  is fixed by  $C$ , if and only if  $\Phi$  is antisymmetric.

The involution  $C$  defines an  $F$ -structure on the  $L$ -vector space  $V^* \otimes_L (V^\iota)^*$  such that the  $F$ -valued points (that is, those fixed by  $C$ ) are exactly the tensors corresponding to antisymmetric sesquilinear forms.

LEMMA 10.1. *If  $V$  is an irreducible representation of  $\Upsilon$ , then the dimension of the  $F$ -vector space of points of  $V^* \otimes_L (V^\iota)^*$  fixed by  $\Upsilon$  and  $C$  is at most one.*

*Proof.* The  $F$ -vector space is a reduction to  $F$  of the  $L$ -vector space of points fixed only by  $\Upsilon$ . Thus, it suffices to prove that the space of points fixed by  $\Upsilon$  has dimension at most one over  $L$ . However, this space is the same as the space of  $\Upsilon$ -equivariant  $L$ -linear maps from  $V^\iota$  to  $V^*$ , and since both are irreducible representations of  $\Upsilon$ , the dimension of the space of maps is at most 1.  $\square$

Fix an embedding  $\sigma : L \hookrightarrow \mathbb{C}$  inducing  $F \hookrightarrow \mathbb{R}$ . Everything in the above discussion can be tensored over  $F$  with  $\mathbb{R}$ . The field  $L$  becomes  $\mathbb{C}$  and  $V \otimes_F \mathbb{R} = V \otimes_L \mathbb{C}$  is a two-dimensional  $\mathbb{C}$ -vector space. We have

$$(V^\iota) \otimes_F \mathbb{R} = \overline{(V \otimes_F \mathbb{R})},$$

and

$$(V^* \otimes_L (V^\iota)^*) \otimes_F \mathbb{R} = (V \otimes_F \mathbb{R})^* \otimes_{\mathbb{C}} \overline{(V \otimes_F \mathbb{R})}^*.$$

The involution  $C$  becomes the involution with a similar formula on this space, whose fixed points are again the tensors corresponding to antisymmetric sesquilinear forms.

The  $\mathbb{R}$ -dimension of the space of points in  $(V^* \otimes_L (V^\iota)^*) \otimes_F \mathbb{R}$  fixed by  $\Upsilon$  and  $C$  is equal to the  $F$ -dimension of the space of fixed points before tensoring. By the above lemma, it is zero or one.

We conclude the following facts from this discussion.

PROPOSITION 10.2. *If for some  $\sigma$  the representation  $V \otimes_{L,\sigma} \mathbb{C}$  admits a  $\Upsilon$ -invariant antisymmetric sesquilinear form  $\Phi_0$ , then the space of such forms is a one-dimensional real vector space. If  $\Phi_0$  exists, then  $V$  admits a  $\Upsilon$ -invariant antisymmetric  $\iota$ -sesquilinear form  $\Phi$  and the space of such forms is again a one-dimensional  $F$ -vector space. For any  $\Phi$  there is a real constant  $\lambda$  such that  $\Phi_0 = \lambda\Phi^\sigma$ .*

### 10.2 Creating a polarized VHS

We go back to the situation considered at the start of this section. We have a rank-two projective  $\mathcal{O}_L$ -module  $P$  with a representation

$$\rho_P : \pi_1(X, x) \rightarrow \mathrm{SL}(P)$$

yielding

$$\rho^\sigma : \pi_1(X, x) \rightarrow \mathrm{SL}(P^\sigma) \cong \mathrm{SL}(2, \mathbb{C})$$

for any embedding  $\sigma : L \rightarrow \mathbb{C}$ . Our original representation  $\rho$  is equal to  $\rho^\eta$ .

Let  $V := P \otimes_{\mathcal{O}_L} L$  be the two-dimensional  $L$ -vector space on which  $\rho_L$  acts. In the notation of the previous subsection  $V \otimes_{L,\sigma} \mathbb{C} = P^\sigma$ . Let  $A_L$  be the local system of  $L$ -vector spaces of rank two corresponding to  $\rho_L$  and for any  $\sigma$  let  $A_\sigma = A_L \otimes_{L,\sigma} \mathbb{C}$  be the local system of complex vector spaces corresponding to the representation  $\rho^\sigma$ .

LEMMA 10.3. *We can choose for each  $\sigma$  a structure of  $\mathbb{C}$ VHS of weight one on the local system  $A_\sigma$  such that*

$$A_{\bar{\sigma}}^{1,0} = \overline{A_\sigma^{0,1}}, \quad A_{\bar{\sigma}}^{0,1} = \overline{A_\sigma^{1,0}}.$$

*Furthermore, these complex variations have hermitian polarizations.*

*Proof.* By hypothesis  $\rho$  is rigid, which implies that each  $\rho^\sigma$  is rigid. Thus, each  $\rho^\sigma$  underlies an hermitian polarized  $\mathbb{C}$ VHS (Theorem 8.1). The fact that we are in rank two means that we can choose these variations to have weight one. If  $\sigma$  is mixed, then the structure has to have weights  $(1, 0)$  and  $(0, 1)$  so it is unique. In this case  $\bar{\sigma}$  is also mixed, and since the complex conjugate variation is one possible choice for  $A_{\bar{\sigma}}$  it is the only choice and we obtain the desired formula.

On the other hand, if  $\sigma$  is positive or negative, then we may choose the structure as we like. In this case  $\bar{\sigma} = \sigma \circ \iota$  has the opposite sign. Thus, we may assume that the complex conjugate choice is made for  $\bar{\sigma}$ , which again gives the formula. □

Fix a choice of variations  $A_\sigma$  as in the lemma. The fact that  $A_\sigma$  has an hermitian polarization implies that there exists a  $\pi_1(X, x)$ -invariant antisymmetric sesquilinear form on  $V \otimes_{L, \sigma} \mathbb{C}$ . By Proposition 10.2, there also exists a  $\pi_1(X, x)$ -invariant antisymmetric  $\iota$ -sesquilinear form  $\Phi$  on  $V$ , and it is unique up to multiplication by scalars in  $F$ .

The remaining problem is to get the right signs for the polarizations on the various pieces. For any embedding  $\sigma : L \rightarrow \mathbb{C}$ , the image  $\Phi^\sigma$  is a real multiple of the polarization form of the variation of Hodge structure  $A_\sigma$  chosen in Lemma 10.3. Thus, there is a sign function depending on  $\Phi$  and denoted

$$g(\Phi, \sigma) \in \{\pm 1\}$$

such that  $g(\Phi, \sigma)\Phi^\sigma$  polarizes the complex variation of Hodge structure  $A_\sigma$ .

LEMMA 10.4. *There is a  $\pi_1(X, x)$ -invariant antisymmetric sesquilinear form  $\Phi$  on  $V$  such that  $g(\Phi, \sigma) = 1$  for all embeddings  $\sigma : L \hookrightarrow \mathbb{C}$ , in other words  $\Phi^\sigma$  polarizes the  $\mathbb{C}$ VHS  $V^\sigma$  for each mixed embedding  $\sigma$ .*

*Proof.* Fix one choice of  $\Phi$  which might not have the right signs. We claim that for each  $\sigma$ ,  $g(\Phi, \sigma \circ \iota) = g(\Phi, \sigma)$ . Recall that we have chosen the variations of Hodge structure  $A_\sigma$  in Lemma 10.3 so that  $A_{\bar{\sigma}} = \overline{A_\sigma}$ . By Lemma 9.1 the form  $\Phi^\sigma$  polarizes  $A_\sigma$  if and only if  $\Phi^{\bar{\sigma}}$  polarizes  $A_{\bar{\sigma}}$ . The same is true if we replace  $\Phi$  by  $-\Phi$ . Thus,  $g(\bar{\sigma}) = g(\sigma)$  as claimed.

For each embedding  $\zeta : F \rightarrow \mathbb{R}$  there are exactly two embeddings  $\sigma, \bar{\sigma} : L \rightarrow \mathbb{C}$  which induce  $\zeta$  on  $F$ . The claim in the previous paragraph says  $g(\Phi, \sigma)$  only depends on  $\zeta = \sigma|_F$ . Abusing notation we write  $g(\Phi, \zeta)$ .

Next let  $\mathbf{sgn}(r) \in \{\pm 1\}$  denote the sign of a non-zero real number. If  $\lambda \in F^*$ , then

$$g(\lambda\Phi, \zeta) = \mathbf{sgn}(\zeta\lambda)g(\Phi, \zeta).$$

Therefore, if we can find  $\lambda \in F^*$  such that  $\mathbf{sgn}(\zeta\lambda) = g(\Phi, \zeta)$  for all  $\zeta : F \rightarrow \mathbb{C}$ , then the form  $\lambda\Phi$  will be a polarization form with the right signs for the lemma.

Now recall that

$$F \otimes_{\mathbb{Q}} \mathbb{R} = \prod_{\zeta : F \rightarrow \mathbb{R}} \mathbb{R}.$$

Since  $F$  is dense in  $F \otimes_{\mathbb{Q}} \mathbb{R}$ , there is a point  $\lambda \in F$  which lies in the open quadrant of  $\prod_{\zeta : F \rightarrow \mathbb{R}} \mathbb{R}$  determined by the specification  $\mathbf{sgn}(\zeta\lambda) = g(\Phi, \zeta)$ . Thus,  $\lambda\Phi$  answers the question. □

Fix a form  $\Phi$  as in the lemma. If  $\sigma$  is a non-mixed embedding then the  $\mathbb{C}$ VHS  $V^\sigma$  is unitary, so it can be chosen as equal to a single Hodge subspace of Hodge type either  $(1, 0)$  or  $(0, 1)$ . We can

choose which one so that  $V^\sigma$  with this Hodge structure is polarized by  $\sqrt{-1}\Phi^\sigma$ . With this condition the choice is unique, and the complex conjugate embedding corresponds to the opposite choice of Hodge type (Lemma 9.1). Thus, if  $\Phi$  is fixed as in the lemma, there is a unique choice of collection of complex variations of Hodge structure for the  $V^\sigma$  such that they are polarized by the  $\sqrt{-1}\Phi^\sigma$  respectively.

We can clear denominators in  $\Phi$  by multiplying it by an appropriate integer. Thus, we may assume that  $\Phi : P \times P \rightarrow \mathcal{O}_L$  is an integral polarization form. We have now obtained the following proposition.

**PROPOSITION 10.5.** *Suppose that  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, \mathbb{C})$  is a rigid, integral Zariski-dense representation. Then there exists: a totally imaginary extension  $L$  of a totally real field  $F$ ; a projective rank-two  $\mathcal{O}_L$ -module  $P$  together with an  $\iota$ -sesquilinear antisymmetric form  $\Phi : P \times P \rightarrow \mathcal{O}_L$ ; a representation  $\rho_P : \pi_1(X, x) \rightarrow \mathcal{SU}(P, \Phi)$ ; and a variation of Hodge structure  $A = \{A_\sigma\}$  of type  $(P, \Phi)$  polarized by  $\Phi$  and whose underlying representation is  $\rho_P$ ; such that for one of the embeddings  $\eta : L \rightarrow \mathbb{C}$  we have  $\rho = \rho^\eta$ .*

By Theorem 9.3 the VHS of type  $(P, \Phi)$  comes from a map  $X \rightarrow \mathfrak{H}(P, \Phi)$  to the Shimura modular stack corresponding to  $(L, P, \Phi)$ .

### 11. Classification

We sum up our main classification results. Suppose that  $X$  is a quasiprojective variety with normal crossings compactification  $\overline{X}$  and  $\overline{X} - X = D$ . Fix a representation  $\rho : \pi_1(X, x) \rightarrow \mathrm{SL}(2, \mathbb{C})$ . Suppose that  $\rho$  is quasiunipotent at infinity, that is, the monodromy elements  $\rho(\gamma_i)$  are quasiunipotent for loops  $\gamma_i$  going around the components  $D_i$  of the divisor at infinity. We also suppose that the image of  $\rho$  is Zariski-dense in  $\mathrm{SL}(2, \mathbb{C})$ .

Recall that we say  $\rho$  is *rigid* if there are no non-trivial global deformations which fix the conjugacy classes of the  $\rho(\gamma_i)$ , see Lemma 6.5, and  $\rho$  is *non-rigid* otherwise.

**THEOREM 11.1.** *Suppose that, with the above hypotheses,  $\rho$  is non-rigid. Then there exists a DM-curve  $Y$ , a rank-two representation  $\rho_Y$  of  $\pi_1(Y)$ , and a map*

$$f : X \rightarrow Y$$

such that  $\rho \cong f^* \rho_Y$ .

*Proof.* This is the contrapositive of Theorem 6.8. □

**THEOREM 11.2.** *Suppose that, with the above hypotheses,  $\rho$  is rigid. Then  $\rho$  is also integral; and there exists a totally imaginary field  $L$  with an embedding  $\eta : L \rightarrow \mathbb{C}$ , a rank-two projective  $\mathcal{O}_L$ -module  $P$  with an antihermitian polarization form  $\Phi$ , and a map to the polydisk Shimura modular stack*

$$f : X \rightarrow \mathfrak{H}(P, \Phi)$$

such that if  $\mathcal{V}^\eta$  is the complex local system on  $\mathfrak{H}(P, \Phi)$  corresponding to the tautological local system tensored with  $\mathbb{C}$  over the embedding  $\eta$ , then  $\rho$  is conjugate to the monodromy representation of  $f^* \mathcal{V}^\eta$ .

*Proof.* By Lemma 7.6,  $\rho$  is integral. Proposition 10.5 gives data  $(L, P, \Phi)$  and a variation of Hodge structure  $V$  of type  $(P, \Phi)$  over  $X$ . If  $\rho_P : \pi_1(X, x) \rightarrow \mathcal{SU}(P, \Phi)$  is the representation underlying this variation of Hodge structure then from Proposition 10.5, for one of the embeddings  $\eta : L \rightarrow \mathbb{C}$  we have  $\rho = \rho_P^\eta$ .

Now Theorems 9.2 and 9.3 say that the variation  $V$  on  $X$  comes from a map  $f : X \rightarrow \mathfrak{H}(P, \Phi)$  in the sense that the pullback of the tautological variation  $f^*(\mathcal{V})$  is the variation  $V$  on  $X$ . The

tautological rank-two complex local systems over  $\mathfrak{H}(P, \Phi)$  all come from the tautological local system of  $\mathcal{O}_L$ -modules of type  $P$  underlying  $\mathcal{V}$ , via the embeddings  $L \hookrightarrow \mathbb{C}$ . Thus, the representation  $\rho \cong \rho_P^\eta$  is the pullback of the tautological rank-two complex local system  $\mathcal{V}^\eta$  on  $\mathfrak{H}(P, \Phi)$ .  $\square$

*Proof of Theorem 2.* Suppose that  $\rho$  is a Zariski-dense representation which is quasiunipotent at infinity. It is either rigid, in which case Theorem 11.2 says that it comes from a polydisk Shimura modular stack, or non-rigid, in which case Theorem 11.1 says that it comes from a DM-curve.  $\square$

There are some cases where a representation on a DM-curve  $Y$  can be rigid, in particular it comes from a map from  $Y$  to a Shimura modular stack. These cases were discussed in § 7.1. In these cases, Theorem 11.2 applies.

**COROLLARY 11.3.** *Any rigid local system of rank two, quasiunipotent at infinity, with Zariski-dense monodromy, is motivic.*

*Proof.* The tautological local systems on Shimura modular stacks are motivic.  $\square$

## 12. Questions

### 12.1 Further development

The most obvious question is: can we give a similar classification, or at least a start, for rank-three representations?

Aside from this, several questions are left open concerning the case of local systems of rank two. For example, what happens in the non-quasiunipotent case? What about local systems with smaller monodromy groups, notably a solvable monodromy group? This question is addressed by Delzant in [Del06, Del07].

For a given DM-curve we would like to have a good understanding of the irreducible components of the moduli spaces of rank-two local systems.

In the present treatment we have restricted to the case where the original  $X$  was a smooth quasiprojective variety. What about the case where  $X$  is a general DM-stack, possibly with singularities?

### 12.2 Calculating the factorization

The transcendental nature of our methods leads naturally to the question of whether one can effectively (and, preferably, easily) calculate the map to a Shimura modular variety corresponding to a given rank-two local system. Concretely,  $V$  might be given in any number of ways:

- (A) by representation matrices corresponding to the generators of the fundamental group (in a Zariski-type calculation of the fundamental group of  $X$ );
- (B) as a vector bundle (probably the trivial bundle, at least if we allow going to a neighborhood in  $X$ ) plus an explicitly given regular–singular connection;
- (C) as a stable parabolic Higgs bundle with vanishing parabolic Chern classes, which might in many cases take on one of several special forms:
  - (C1) a stable vector bundle with  $\theta = 0$  over a projective variety;
  - (C1′) a stable parabolic vector bundle with  $\theta = 0$ ;
  - (C2) a stable system of Hodge bundles, which in the rank-two case and excepting (C1) above means a pair of line bundles and a map

$$\theta : L \rightarrow M \otimes \Omega_X^1$$

- (C2′) a stable system of Hodge bundles with parabolic structure and/or logarithmic  $\theta$ .

For each kind of concrete example of  $V$  there could arise the question of how to explicitly construct the map from  $X$  (or a finite covering space  $X'$  of  $X$ ) to a Shimura modular variety plus an isomorphism between our given object and the pullback of the universal object.

While it might be possible to prove using some general principles that there exists an algorithm which solves this problem in principle, we would actually like to have a good method which works on specific examples and which is related to geometry [ACT02, AB00, Dim06, DN01, FM97, KM98, RT97, Zuo94].

There are now known some specific examples. For example Panov constructs rank-two unitary representations corresponding to polyhedral Kähler structures on some surfaces in his thesis [Pan04].

Any algebraic solution of the Painlevé VI equation will correspond to a rank-two local system on a surface. Such solutions have been constructed by Hitchin [Hit95], Dubrovin [Dub96], Boalch [Boa05], Ben Hamed-Gavrilov [BG05]. In many cases the solutions were shown to have geometric origin already.

More generally there are various potential sources of other examples [BK92, Cam04, Del86, DN01, Moi77, Nor83, SU86, Tol90], in particular line or hyperplane arrangements as was the case in Panov's construction [ACT02, CO00, Fy07, MT88, PS02]. These might in the future lead to constructions of rank-two local systems for which the classification results would apply.

To what extent is the factorization map uniquely determined by the representation? This is easy for factorization through an orbicurve, but might be an interesting question for factorization through a Shimura modular stack. What kind of Shimura modular stack is associated to a given local system? For example, what is its dimension? Are there details specific to the theory of Hilbert modular varieties that could be extended to the polydisk Shimura case? On the other side, how do we tell, given  $(P, \Phi)$ , how many mixed embeddings there are?

The problem of calculating explicitly the factorization is also interesting when  $X$  is a curve. In § 5.3 we have started to look at the real geometry of the pluriharmonic map this case. This was very preliminary but it looks interesting and should be further pursued. How does the representation  $\rho : \pi_1 \rightarrow \mathrm{SL}(2, K)$  determine the geometry of the map from  $\tilde{X}$  to the tree?

The analytic proof that non-rigid representations factor, which we have sidestepped here, also leads to an effectivity question. The space of representations whose Higgs fields are nilpotent is compact—can we give an explicit bound for the size of these representations?

## REFERENCES

- AV02 D. Abramovich and A. Vistoli, *Compactifying the space of stable maps*, J. Amer. Math. Soc. **15** (2002), 27–75.
- ACT02 D. Allcock, J. Carlson and D. Toledo, *Orthogonal complex hyperplane arrangements*, in *Symposium in Honor of C. H. Clemens (Salt Lake City, 2000)*, Contemporary Mathematics, vol. 312 (American Mathematical Society, Providence, RI, 2002), 1–8.
- AB00 J. Amorós and I. Bauer, *On the number of defining relations for nonfibered Kähler groups*, Internat. J. Math. **11** (2000), 285–290.
- ABCKT96 J. Amorós, M. Burger, K. Corlette, D. Kotschick and D. Toledo, *Fundamental groups of compact Kähler manifolds*, AMS Mathematical Surveys and Monographs, vol. 44 (American Mathematical Society, Providence, RI, 1996).
- Ara97 D. Arapura, *Geometry of cohomology support loci for local systems. I*, J. Algebraic Geom. **6** (1997), 563–597.
- ABR92 D. Arapura, P. Bressler and M. Ramachandran, *On the fundamental group of a compact Kähler manifold*, Duke Math. J. **68** (1992), 477–488.

- BBN01 V. Balaji, I. Biswas and D. Nagaraj, *Principal bundles over projective manifolds with parabolic structure over a divisor*, Tohoku Math. J. (2) **53** (2001), 337–367.
- Bas80 H. Bass, *Groups of integral representation type*, Pacific J. Math. **86** (1980), 15–51.
- Bea92 A. Beauville, *Annulation du  $H^1$  pour les fibrés en droites plats*, in *Complex algebraic varieties (Bayreuth, 1990)*, Lecture Notes in Mathematics, vol. 1507 (Springer, Berlin, 1992), 1–15.
- BN06 K. Behrend and B. Noohi, *Uniformization of Deligne–Mumford curves*, J. reine angew. Math. **599** (2006), 111–153.
- Bel82 S. Bell, *The Bergman kernel function and proper holomorphic mappings*, Trans. Amer. Math. Soc. **270** (1982), 685–691.
- BG05 B. Ben Hamed and L. Gavrilov, *Families of Painlevé VI equations having a common solution*, Int. Math. Res. Not. **60** (2005), 3727–3752.
- Biq96 O. Biquard, *Sur les fibrés paraboliques sur une surface complexe*, J. London Math. Soc. (2) **53** (1996), 302–316.
- Bis97 I. Biswas, *Parabolic bundles as orbifold bundles*, Duke Math. J. **88** (1997), 305–325.
- Boa05 P. Boalch, *From Klein to Painlevé via Fourier, Laplace and Jimbo*, Proc. London Math. Soc. (3) **90** (2005), 167–208.
- BY96 H. Boden and K. Yokogawa, *Moduli spaces of parabolic Higgs bundles and parabolic  $K(D)$  pairs over smooth curves, I*, Internat. J. Math. **7** (1996), 573–598.
- BK92 F. Bogomolov and L. Katzarkov, *Complex projective surfaces and infinite groups*, Geom. Funct. Anal. **8** (1998), 243–272.
- Bre94 L. Breen, *On the classification of 2-gerbes and 2-stacks*, Astérisque, vol. 225 (Société Mathématique de France, Paris, 1994).
- Cam94 F. Campana, *Remarques sur le revêtement universel des variétés kählériennes compactes*, Bull. Soc. Math. France **122** (1994), 255–284.
- Cam04 F. Campana, *Orbifolds, special varieties and classification theory*, Ann. Inst. Fourier (Grenoble) **54** (2004), 499–630.
- CSSB04 V. Caselles, G. Sapiro, A. Solé and C. Ballester, *Morse description and morphological encoding of continuous data*, Multiscale Model. Simul. **2** (2004), 179–209.
- Cat96 F. Catanese, *Fundamental groups with few relations*, in *Higher dimensional complex varieties (Trento 1994)* (De Gruyter, Berlin, 1996), 163–165.
- CKS87 E. Cattani, A. Kaplan and W. Schmid,  *$L^2$  and intersection cohomologies for a polarizable variation of Hodge structure*, Invent. Math. **87** (1987), 217–252.
- CO00 D. Cohen and P. Orlik, *Arrangements and local systems*, Math. Res. Lett. **7** (2000), 299–316.
- CEHNP04 K. Cole-McLaughlin, H. Edelsbrunner, J. Harer, V. Natarajan and V. Pascucci, *Loops in Reeb graphs of 2-manifolds*, Discrete Comput. Geom. **32** (2004), 231–244.
- Cor88 K. Corlette, *Flat  $G$ -bundles with canonical metrics*, J. Differential Geom. **28** (1988), 361–382.
- Cor91 K. Corlette, *Rigid representations of Kählerian fundamental groups*, J. Differential Geom. **33** (1991), 239–252.
- Car92 K. Corlette, *Nonabelian Hodge theory and square integrability*, in *Actes du Colloque “Géométrie et Topologie des Variétés Projectives (Toulouse 1992)”*, Preprint (1992), Laboratoire de Topologie et Géométrie, Université Toulouse III.
- CS83 M. Culler and P. Shalen, *Varieties of group representations and splittings of 3-manifolds*, Ann. of Math. (2) **117** (1983), 109–146.
- Dar00 H. Darmon, *Rigid local systems, Hilbert modular forms, and Fermat’s last theorem*, Duke Math. J. **102** (2000), 413–449.
- Del71 P. Deligne, *Travaux de Shimura*, in *Séminaire Bourbaki No. 389 (1970/71)*, Lecture Notes in Mathematics, vol. 244 (Springer, Berlin, 1971), 123–165.
- Del79 P. Deligne, *Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques*, in *Automorphic forms, representations and  $L$ -functions (Corvallis, 1977)*,



- Proceedings of Symposia in Pure Mathematics, vol. 23 (American Mathematical Society, Providence, RI, 1979), 247–289.
- Del86 P. Deligne and G. Mostow, *Monodromy of hypergeometric functions and nonlattice integral monodromy*, Publ. Math. Inst. Hautes Études Sci. **63** (1986), 5–89.
- DM69 P. Deligne and D. Mumford, *The irreducibility of the space of curves of given genus*, Publ. Math. Inst. Hautes Études Sci. **36** (1969), 75–109.
- Del06 T. Delzant, *L'invariant de Bieri Neumann Strebel des groupes fondamentaux des variétés kähleriennes*, Preprint (2006), math.DG/0603038, Math. Ann., to appear.
- Del07 T. Delzant, *Trees, valuations and the Green–Lazarsfeld set*, Preprint (2007), math.AG/0702477, Geom. Funct. Anal., to appear.
- DR00 M. Dettweiler and S. Reiter, *An algorithm of Katz and its application to the inverse Galois problem*, *Algorithmic methods in Galois theory*, J. Symbolic Comput. **30** (2000), 761–798.
- DO85 K. Diederich and T. Ohsawa, *Harmonic mappings and disc bundles over compact Kähler manifolds*, Publ. Res. Inst. Math. Sci. **21** (1985), 819–833.
- Dim06 A. Dimca, *Pencils of plane curves and characteristic varieties*, Preprint (2006), math.AG/0606442.
- DN01 A. Dimca and A. Nemethi, *On the monodromy of complex polynomials*, Duke Math. J. **108** (2001), 199–209.
- Don85 S. Donaldson, *Anti self-dual Yang–Mills connections over complex algebraic surfaces and stable vector bundles*, Proc. London Math. Soc. (3) **50** (1985), 1–26.
- Don87 S. Donaldson, *Twisted harmonic maps and the self-duality equations*, Proc. London Math. Soc. (3) **55** (1987), 127–131.
- Dub96 B. Dubrovin, *Geometry of 2D topological field theories*, in *Integrable systems and quantum groups (Montecatini Terme, 1993)*, Lecture Notes in Mathematics, vol. 1620 (Springer, Berlin, 1996), 120–348.
- ES64 J. Eells and J. Sampson, *Harmonic mappings of Riemannian manifolds*, Amer. J. Math. **86** (1964), 109–160.
- EP03 M. Entov and L. Polterovich, *Calabi quasimorphism and quantum homology*, Int. Math. Res. Not. **30** (2003), 1635–1676.
- Eys04 P. Eyssidieux, *Sur la convexité holomorphe des revêtements linéaires réductifs d'une variété projective algébrique complexe*, Invent. Math. **156** (2004), 503–564.
- Fy07 M. Falk and S. Yuzvinsky, *Multinets, resonance varieties, and pencils of plane curves*, Compositio Math. **143** (2007), 1069–1088.
- FM97 R. Friedman and J. Morgan, *Algebraic surfaces and Seiberg–Witten invariants*, J. Algebraic Geom. **6** (1997), 445–479.
- Fre90 E. Freitag, *Hilbert modular forms* (Springer, Berlin, 1990).
- GKM08 T. Gelander, A. Karlsson and G. Margulis, *Superrigidity, generalized harmonic maps and uniformly convex spaces*, Geom. Funct. Anal. **17** (2008), 1524–1550.
- Gor02 E. Goren, *Lectures on Hilbert modular varieties and modular forms (with the assistance of Marc-Hubert Nicole)*, CRM Monograph Series, vol. 14 (American Mathematical Society, Providence, RI, 2002).
- GG03 M. Green and P. Griffiths, *Hodge-theoretic invariants for algebraic cycles*, Int. Math. Res. Not. **9** (2003), 477–510.
- Gro89 M. Gromov, *Sur le groupe fondamental d'une variété kählérienne*, C. R. Acad. Sci. Paris Sér. I Math. **308** (1989), 67–70.
- GS92 M. Gromov and R. Schoen, *Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one*, Publ. Math. Inst. Hautes Études Sci. **76** (1992), 165–246.
- Hai06 R. Hain, *Finiteness and Torelli spaces*, in *Problems on mapping class groups and related topics*, Proceedings Symposia in Pure Mathematics, vol. 74, (American Mathematical Society, Providence, RI, 2006), 57–70.

- Hit95 N. Hitchin, *Poncelet polygons and the Painlevé equations*, in *Geometry and analysis (Bombay, 1992)* (Tata Institute of Fundamental Research, Bombay, 1995), 151–185.
- HU74 F. Hirzebruch and A. Van de Ven, *Hilbert modular surfaces and the classification of algebraic surfaces*, *Invent. Math.* **23** (1974), 1–29.
- IIS06 M. Inaba, K. Iwasaki and M.-H. Saito, *Moduli of stable parabolic connections, Riemann–Hilbert correspondence and geometry of Painlevé equation of type VI, Part I*, *Publ. Res. Inst. Math. Sci.* **42** (2006), 987–1089.
- JY86 J. Jost and S.-T. Yau, *The strong rigidity of locally symmetric complex manifolds of rank one and finite volume*, *Math. Ann.* **275** (1986), 291–304.
- JZ96 J. Jost and K. Zuo, *Harmonic maps and  $SL(r, \mathbb{C})$  representations of fundamental groups of quasi-projective manifolds*, *J. Algebraic Geom.* **5** (1996), 77–106.
- JZ00 J. Jost and K. Zuo, *Harmonic maps into Bruhat-Tits buildings and factorizations of  $p$ -adically unbounded representations of  $\pi_1$  of algebraic varieties I*, *J. Algebraic Geom.* **9** (2000), 1–42.
- KM98 M. Kapovich and J. Millson, *On representation varieties of Artin groups, projective arrangements and the fundamental groups of smooth projective varieties*, *Publ. Math. Inst. Hautes Études Sci.* **88** (1998), 5–95.
- Kat90 N. Katz, *Exponential sums and differential equations*, *Annals of Mathematics Studies*, vol. 124 (Princeton University Press, Princeton, NJ, 1990).
- Kat96 N. Katz, *Rigid local systems*, *Annals of Mathematics Studies*, vol. 139 (Princeton University Press, Princeton, NJ, 1996).
- Kat94 L. Katzarkov, *Factorization theorems for the representations of the fundamental groups of quasiprojective varieties and some applications*, Preprint (1994), alg-geom/9402012v1.
- Kat97 L. Katzarkov, *On the Shafarevich maps*, in *Algebraic geometry (Santa Cruz, 1995)*, *Proceedings of Symposia in Pure Mathematics*, vol. 62(2) (American Mathematical Society, Providence, RI, 1997), 173–216.
- KR98 L. Katzarkov and M. Ramachandran, *On the universal coverings of algebraic surfaces*, *Ann. Sci. École Norm. Sup.* **31** (1998), 525–535.
- Kaw78 T. Kawasaki, *The signature theorem for  $V$ -manifolds*, *Topology* **17** (1978), 75–83.
- KM97 S. Keel and S. Mori, *Quotients by groupoids*, *Ann. of Math. (2)* **145** (1997), 193–213.
- Kli03 B. Klingler, *Sur la rigidité de certains groupes fondamentaux, l’arithmétique des réseaux hyperboliques complexes, et les “faux plans projectifs”*, *Invent. Math.* **153** (2003), 105–143.
- KO75 S. Kobayashi and T. Ochiai, *Meromorphic mappings onto compact complex spaces of general type*, *Invent. Math.* **31** (1975), 7–16.
- Kol93 J. Kollar, *Shafarevich maps and plurigenera of algebraic varieties*, *Invent. Math.* **113** (1993), 177–215.
- Kon93 H. Konno, *Construction of the moduli space of stable parabolic Higgs bundles on a Riemann surface*, *J. Math. Soc. Japan* **45** (1993), 253–276.
- Kon07 M. Kontsevich, *Notes on motives in finite characteristic*, Preprint (2007), math.AG/0702206.
- Kot92 R. Kottwitz, *Points on some Shimura varieties over finite fields*, *J. Amer. Math. Soc.* **5** (1992), 373–444.
- Lan62 S. Lang, *Diophantine geometry*, *Interscience Tracts in Pure and Applied Mathematics*, vol. 11 (Wiley, New York, 1962).
- LO06 Y. Laszlo and M. Olsson, *The six operations for sheaves on Artin stacks II: adic coefficients*, Preprint (2006), math.AG/0603680.
- LP92 M. Larsen and R. Pink, *On  $l$ -independence of algebraic monodromy groups in compatible systems of representations*, *Invent. Math.* **107** (1992), 603–636.
- Lar92 M. Larsen, *Arithmetic compactification of some Shimura surfaces*, in *The zeta functions of Picard modular surfaces* (Université Montréal, Montréal, 1992), 31–45.
- LM00 G. Laumon and L. Moret-Bailly, *Champs algébriques*, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 39 (Springer, Berlin, 2000).

- Li00 J. Li, *Hermitian–Einstein metrics and Chern number inequalities on parabolic stable bundles over Kähler manifolds*, *Comm. Anal. Geom.* **8** (2000), 445–475.
- LN99 J. Li and M. S. Narasimhan, *Hermitian–Einstein metrics on parabolic stable bundles*, *Acta Math. Sin. (Engl. Ser.)* **15** (1999), 93–114.
- LN01 J. Li and M. S. Narasimhan, *A note on Hermitian–Einstein metrics on parabolic stable bundles*, *Acta Math. Sin. (Engl. Ser.)* **17** (2001), 77–80.
- LM85 A. Lubotsky and A. Magid, *Varieties of representations of finitely generated groups*, *Mem. Amer. Math. Soc.*, vol. 58 (American Mathematical Society, Providence, RI, 1985).
- Mcm03 C. McMullen, *Billiards and Teichmüller curves on Hilbert modular surfaces*, *J. Amer. Math. Soc.* **16** (2003), 857–885.
- Moc02 T. Mochizuki, *Asymptotic behaviour of tame nilpotent harmonic bundles with trivial parabolic structure*, *J. Differential Geom.* **62** (2002), 351–559.
- Moc07 T. Mochizuki, *Asymptotic behaviour of tame harmonic bundles and an application to pure twistor  $D$ -modules*, *Mem. Amer. Math. Soc.*, vol. 185, nos 869–870 (American Mathematical Society Providence RI, 2007).
- Moi77 B. Moishezon, *Complex surfaces and connected sums of complex projective planes*, appendix by R. Livne, *Lecture Notes in Mathematics*, vol. 603 (Springer, Berlin, 1977).
- MT88 B. Moishezon and M. Teicher, *Braid group technique in complex geometry. I. Line arrangements in  $\mathbb{C}P^2$ , Braids (Santa Cruz, 1986)*, *Contemporary Mathematics*, vol. 78 (American Mathematical Society, Providence, RI, 1988), 425–555.
- Nak96 H. Nakajima, *Hyper-Kähler structures on moduli spaces of parabolic Higgs bundles on Riemann surfaces, Moduli of vector bundles (Sanda, Kyoto, 1994)*, *Lecture Notes in Pure and Applied Mathematics*, vol. 179 (Chapman & Hall/CRL, Boca Raton, FL, 1996), 199–208.
- NR01a T. Napier and M. Ramachandran, *Hyperbolic Kähler manifolds and proper holomorphic mappings to Riemann surfaces*, *Geom. Funct. Anal.* **11** (2001a), 382–406.
- NR06 T. Napier and M. Ramachandran, *Thompson’s group  $F$  is not Kähler*, in *Topological and asymptotic aspects of group theory*, *Contemporary Mathematics*, vol. 394, (American Mathematical Society, Providence, RI, 2006), 197–201.
- NR08 T. Napier and M. Ramachandran, *Filtered ends, proper holomorphic mappings of Kähler manifolds to Riemann surfaces, and Kähler groups*, *Geom. Funct. Anal.* **17** (2008), 1621–1654.
- NR01b L. Ni and H. Ren, *Hermitian–Einstein metrics for vector bundles on complete Kähler manifolds*, *Trans. Amer. Math. Soc.* **353** (2001b), 441–456.
- Nit93 N. Nitsure, *Moduli of semistable logarithmic connections*, *J. Amer. Math. Soc.* **6** (1993), 597–609.
- Noo04 B. Noohi, *Fundamental groups of algebraic stacks*, *J. Inst. Math. Jussieu* **3** (2004), 69–103.
- Nor83 M. Nori, *Zariski’s conjecture and related problems*, *Ann. Sci. École Norm. Sup.* **16** (1983), 305–344.
- Pan04 D. Panov, *Polyhedral Kähler Manifolds*, Thesis (2004), École Polytechnique.
- PS02 S. Papadima and A. Suciu, *Higher homotopy groups of complements of complex hyperplane arrangements*, *Adv. Math.* **165** (2002), 71–100.
- Pap00 G. Pappas, *On the arithmetic moduli schemes of PEL Shimura varieties*, *J. Algebraic Geom.* **9** (2000), 577–605.
- Pro76 C. Procesi, *The invariant theory of  $n \times n$  matrices*, *Adv. Math.* **19** (1976), 306–381.
- Rap78 M. Rapoport, *Compactifications de l’espace de modules de Hilbert–Blumenthal*, *Compositio Math.* **36** (1978), 255–335.
- RZ96 M. Rapoport and T. Zink, *Period spaces for  $p$ -divisible groups*, *Annals of Mathematics Studies*, vol. 141 (Princeton University Press, Princeton, NJ, 1996).
- Ree46 G. Reeb, *Sur les points singuliers d’une forme de Pfaff complètement intégrable ou d’une fonction numérique*, *C. R. Acad. Sci. Paris* **222** (1946), 847–849.

- Rez02 A. Reznikov, *The structure of Kähler groups. I. Second cohomology*, in *Motives, polylogarithms and Hodge theory (Irvine, 1998)*, International Press Lecture Series, vol. 3(II) (International Press, Somerville, MA, 2002), 717–730.
- RT97 A. Robb and M. Teicher, *Applications of braid group techniques to the decomposition of moduli spaces, new examples*, (*Special issue on braid groups and related topics, Jerusalem, 1995*), *Topology Appl.* **78** (1997), 143–151.
- Sat56 I. Satake, *On a generalization of the notion of manifold*, *Proc. Natl. Acad. Sci. USA* **42** (1956), 359–363.
- Sat57 I. Satake, *The Gauss–Bonnet theorem for  $V$ -manifolds*, *J. Math. Soc. Japan* **9** (1957), 464–492.
- Ser77 J.-P. Serre, *Arbres, amalgames,  $SL_2$* , in *Résumé avec la collaboration de Hyman Bass*, Astérisque, vol. 46 (Société Mathématique de France, Paris, 1977).
- Ser80 J.-P. Serre, *Trees*, *Monographs in Mathematics* (Springer, Berlin, 1980).
- Shi86 B. Shiffman, *Complete characterization of holomorphic chains of codimension one*, *Math. Ann.* **274** (1986), 233–256.
- Shi94 B. Shiffman, *Separately meromorphic mappings into compact Kähler manifolds*, in *Contributions to complex analysis and analytic geometry*, *Aspects of Mathematics*, vol. E26 (Vieweg, Wiesbaden, 1994), 243–250.
- Sic69 J. Siciak, *Separately analytic functions and envelopes of holomorphy of some lower dimensional subsets of  $\mathbb{C}^n$* , *Ann. Polon. Math.* **22** (1969/1970), 145–171.
- Sim90 C. Simpson, *Harmonic bundles on noncompact curves*, *J. Amer. Math. Soc.* **3** (1990), 713–770.
- Sim91 C. Simpson, *The ubiquity of variations of Hodge structure*, in *Complex geometry and Lie theory (Sundance, 1989)*, *Proceedings of Symposia in Pure Mathematics*, vol. 53 (American Mathematical Society, Providence, RI, 1991), 329–348.
- Sim92 C. Simpson, *Higgs bundles and local systems*, *Publ. Math. Inst. Hautes Études Sci.* **75** (1992), 5–95.
- Sim93 C. Simpson, *Lefschetz theorems for the integral leaves of a holomorphic one-form*, *Compositio Math.* **87** (1993), 99–113.
- Sim95 C. Simpson, *Moduli of representations of the fundamental group of a smooth projective variety, II*, *Publ. Math. Inst. Hautes Études Sci.* **80** (1995), 5–79.
- Siu80 Y.-T. Siu, *The complex analyticity of harmonic maps and the strong rigidity of compact Kähler manifolds*, *Ann. of Math. (2)* **112** (1980), 73–111.
- SY82 Y.-T. Siu and S.-T. Yau, *Compactification of negatively curved complete Kähler manifolds of finite volume*, *Seminar on Differential Geometry*, *Annals of Mathematics Studies*, vol. 102 (Princeton University Press, Princeton, NJ, 1982) 363–380.
- SU86 A. Sommese and A. Van de Ven, *Homotopy groups of pullbacks of varieties*, *Nagoya Math. J.* **102** (1986), 79–90.
- SW01 B. Steer and A. Wren, *The Donaldson–Hitchin–Kobayashi correspondence for parabolic bundles over orbifold surfaces*, *Canad. J. Math.* **53** (2001), 1309–1339.
- Ste06 D. Stepanov, *A remark on the dual complex of a resolution of singularities*, *Russian Math. Surveys* **61** (2006), 181–183.
- Tha02 M. Thaddeus, *Variation of moduli of parabolic Higgs bundles*, *J. reine angew. Math.* **547** (2002), 1–14.
- Thu07 A. Thuillier, *Géométrie toroidale et géométrie analytique non archimédienne. Application au type d’homotopie de certains schémas formels*, *Manuscripta Math.* **123** (2007), 381–451.
- Toe99 B. Toen,  *$K$ -theory and cohomology of algebraic stacks: Riemann–Roch theorems,  $D$ -modules and GAGA theorems*, *Doctoral dissertation* (1999), Toulouse, math.AG/9908097.
- Tol90 D. Toledo, *Examples of fundamental groups of compact Kähler manifolds*, *Bull. London Math. Soc.* **22** (1990), 339–343.
- Van88 G. van der Geer, *Hilbert modular surfaces*, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 16 (Springer, Berlin, 1988).

- Vie05 E. Viehweg and K. Zuo, *Special subvarieties of  $A_g$* , Preprint (2005), math.AG/0509037, Proc. of the ICCM 2004, to appear.
- Vis89 A. Vistoli, *Intersection theory on algebraic stacks and on their moduli spaces*, Invent. Math. **97** (1989), 613–670.
- Wed99 T. Wedhorn, *Ordinariness in good reductions of Shimura varieties of PEL-type*, Ann. Sci. École Norm. Sup. **32** (1999), 575–618.
- Yok93 K. Yokogawa, *Compactification of moduli of parabolic sheaves and moduli of parabolic Higgs sheaves*, J. Math. Kyoto Univ. **33** (1993), 451–504.
- Zuo94 K. Zuo, *Factorizations of nonrigid Zariski dense representations of  $\pi_1$  of projective algebraic manifolds*, Invent Math. **118** (1994), 37–46.

Kevin Corlette kevin@math.uchicago.edu

University of Chicago, 5734 University Avenue, Chicago, IL 60637, USA

Carlos Simpson carlos@unice.fr

CNRS, Laboratoire J. A. Dieudonné, UMR 6621, Université de Nice-Sophia Antipolis, 06108 Nice, Cedex 2, France