Linearity properties of Shimura varieties, I

Ben Moonen

Introduction

0.1 Let $Sh_K(G,\mathfrak{X})$ be a Shimura variety. We call an irreducible subvariety $Z \hookrightarrow Sh_K(G,\mathfrak{X})$ a "subvariety of Hodge type" if, up to a Hecke correspondence, Z is an irreducible component of a Shimura subvariety—see Section 2 for a more precise definition. The Deligne formalism of Shimura varieties provides a clean, almost purely group theoretical, description of the Shimura subvarieties of $Sh_K(G,\mathfrak{X})$. In some situations, however, it remains very difficult to describe which subvarieties are of Hodge type. We will illustrate this with some examples in 0.3 below.

The central theme of this paper is that we try to find criteria for an algebraic subvariety Z of a Shimura variety to be a subvariety of Hodge type. The general shape of our results is that subvarieties of Hodge type can be characterized by certain "linearity properties", to be tested at one single point. In this first part of the paper we work with an arbitrary Shimura variety $Sh_K(G,\mathfrak{X})$ over $\mathbb{C}$. Here the linearity property of interest is that of being "totally geodesic" (see 4.1). We give a complete description of totally geodesic algebraic subvarieties $Z \hookrightarrow Sh_K(G,\mathfrak{X})$, and in particular we prove that an algebraic subvariety Z is of Hodge type if and only if it is totally geodesic and it contains at least one special point.

Next we discuss in some detail an example which demonstrates that a Shimura variety may contain totally geodesic algebraic subvarieties which are not of Hodge type (i.e., which do not contain any special points). We show that this example, which we think is very instructive, provides a negative answer to two problems posed in André's book [1].

Another spin-off is that the example makes clear, in very geometrical terms, how non-rigid families of abelian varieties arise. Here we recall that, writing $\mathsf{A}_{g,1,n}$ for the moduli space of principally polarized g-dimensional abelian varieties with a level n structure, we can describe $\mathsf{A}_{g,1,n} \otimes \mathbb{C}$ as a Shimura variety associated to the group $\mathrm{CSp}_{2g,\mathbb{Q}}$. In its simplest form the idea describing non-rigid families of abelian varieties is that, inside some $\mathsf{A}_{g,1,n} \otimes \mathbb{C}$, we have a Shimura variety S of Hodge type which, as a variety, is the product $S_1 \times S_2$ of two Shimura varieties (which themselves are *not* of Hodge type). Non-rigid families of abelian varieties are then obtained by looking at subvarieties of the form $S_1 \times \{a\}$, where deformations are given by "moving the point $a \in S_2$". Notice that the decomposition of S as a product is due to the fact that its adjoint Shimura variety is a product—see Section 3 for more on this. The surprising fact, proved in Section 6, is that non-rigidity of abelian schemes always arises from such a product decomposition (on the adjoint level) of a Shimura variety. For a precise statement we refer to Theorem 6.4.

0.2 In Part II of this paper we study subvarieties of Hodge type in mixed characteristics, and in particular their local structure at the ordinary locus in characteristic p. Our main result in this context is that subvarieties of Hodge type are characterized by a certain "formal linearity property", i.e., linearity with respect to Serre-Tate coordinates.

To make the analogy between the two characterizations even clearer, we introduce in Section 5 a "Serre-Tate group structure" over $\mathbb{C}$, and we re-interpret the total geodesicness of an algebraic subvariety $Z \hookrightarrow Sh_K(G, \mathfrak{X})$ in terms of this formal group structure. In this way, we arrive at a uniform formulation of our characterizations over $\mathbb{C}$ and in mixed characteristics, respectively.

Our proof that linearity w.r.t. Serre-Tate coordinates (over $\mathbb{C}$) is equivalent (for algebraic varieties) to the property of being totally geodesic, makes essential use of a monodromy argument. We consider a component $\widetilde{Z}$ of the preimage of Z under a uniformization map $X \subset \mathfrak{X} \to Sh_K(G, \mathfrak{X})$. If Z is linear w.r.t. Serre-Tate coordinates at some point y, then we obtain very precise information about the equations defining $\widetilde{Z} \hookrightarrow X$ locally at a point x above y. A monodromy argument, combined with a result of Y. André, then allows us to show that there is a "sufficiently big" algebraic subgroup $H \subset G$ such that $\widetilde{Z}$ is stable under the action of $H(\mathbb{R})^+$ on X, from which we deduce that Z is totally geodesic.

0.3 To conclude this introduction, let us mention two problems that have motivated our research. These problems should explain why it is of interest to have a direct characterization of subvarieties of Hodge type.

Conjecture. (Coleman, cf. [12]) For a fixed $g \geq 4$, there are finitely many smooth projective genus g curves C over $\mathbb{C}$ (taken up to isomorphism) such that $\mathrm{Jac}(C)$ is of CM-type.

As a matter of fact, the conjecture is false for $g = 4$ and $g = 6$. The reason for this is that for these genera one can find subvarieties of Hodge type which are contained in the Torelli locus—see the paper [13] by A.J. de Jong and R. Noot. For $g = 5$ and $g \geq 7$ the Coleman conjecture remains, to our knowledge, at present completely open.

Conjecture. (Oort, cf. [24]) Let $Z \hookrightarrow \mathsf{A}_{g,1,n} \otimes \mathbb{C}$ be an irreducible algebraic subvariety such that the CM-points on Z are dense for the Zariski topology. Then Z is a subvariety of Hodge type.

This conjecture will be discussed in more detail in the second part of this paper, where we prove that it is equivalent to a certain statement about the reduction behaviour of the CM-points on the variety Z in question, and where we prove the conjecture in a special situation. Very recently, Y. André has obtained a proof of Oort's conjecture for subvarieties of $\mathsf{A}_1 \times \mathsf{A}_1$.

Oort's conjecture can be viewed as a first step towards the Coleman conjecture—the second step then would be to decide wether there exist subvarieties of Hodge type which are contained in the Torelli locus. In this connection, let us mention that R. Hain recently obtained some results, giving restrictions on the possible subvarieties of Hodge type that can be contained in the (open) Torelli locus.

Acknowledgements. This paper contains the results of the second chapter of my Utrecht Ph.D. thesis. It is a great pleasure to thank my advisors, F. Oort and A.J. de Jong, for their constant support, their friendship, and for everything they have taught me.

I wish to thank Harvard University for its hospitality during my stay as a Visiting Fellow in the first half of 1994, when part of the research for this paper was done. Also I wish to thank all those who have shown interest in my work—in particular I thank B. Edixhoven, D. Kazhdan, R. Pink, M.-H. Saito and T. Zink for their valuable comments. Finally, I thank the referee for several useful suggestions.

§1 Hodge-theoretical preliminaries

1.1 For the basic definitions from Hodge theory we refer to [16], especially Sect. 2. Set $\mathbb{S} := \mathrm{Res}_{\mathbb{C}/\mathbb{R}}\,\mathbb{G}_{\mathrm{m}}$, and write $w\colon \mathbb{G}_{\mathrm{m},\mathbb{R}} \to \mathbb{S}$ for the cocharacter which on real-valued points is given by $\mathbb{R}^* \subset \mathbb{C}^*$. As explained in loc. cit., a polarizable pure $\mathbb{Q}$-Hodge structure of weight n with underlying $\mathbb{Q}$-vector space V can be described by giving a homomorphism of algebraic groups

$$h\colon \mathbb{S} \to \mathrm{GL}(V)_{\mathbb{R}}\,,$$

such that $h \circ w$ is given by $\mathbb{C}^* = \mathbb{S}(\mathbb{R}) \ni z \mapsto z^{-n} \cdot \mathrm{Id}_V$. (Sign conventions as in [17], i.e., *opposite* to [16, (2.1.5.1)].) We define the Mumford-Tate group $\mathrm{MT}(V)$ as the smallest algebraic subgroup of $\mathrm{GL}(V)$ such that h factors through $\mathrm{MT}(V)_{\mathbb{R}}$.

Fix an element $i \in \mathbb{C}$ with $i^2 = -1$, and use this to identify the $\mathbb{Z}$-modules $\mathbb{Z}$ and $\mathbb{Z}(n)$ as in [16, (2.1.14)]. Via this identification a polarization $\psi\colon V^{\otimes 2} \to \mathbb{Q}(-n)$ gives a bilinear form on V, which we again call ψ. This form is symmetric if n is even, skew-symmetric if n is odd, and $\mathrm{MT}(V)$ is a reductive subgroup of the group of elements $g \in \mathrm{GL}(V)$ which preserve ψ up to a scalar.

The Mumford-Tate group as defined here is the image of the Mumford-Tate group $\mathrm{MT}'(V) \subseteq \mathrm{GL}(V) \times \mathbb{G}_{\mathrm{m}}$ defined in [18] (called the extended Mumford-Tate group in [23]) under the projection to $\mathrm{GL}(V)$. If the weight n is non-zero then the projection $\mathrm{MT}'(V) \twoheadrightarrow \mathrm{MT}(V)$ is an isogeny; for $n = 0$ it has $\{1\} \times \mathbb{G}_{\mathrm{m}}$ as its kernel.

1.2 Let S be a connected complex manifold. Recall (see [9, Sect. 2] for example) that a polarized variation of $\mathbb{Q}$-Hodge structure (abbreviated VHS) of weight n over S is a triplet $\mathcal{V} = (\mathcal{V}_{\mathbb{Q}}, \mathcal{F}^{\bullet}, Q)$, where $\mathcal{V}_{\mathbb{Q}}$ is a local system of finite-dimensional $\mathbb{Q}$-vector spaces, $\mathcal{F}^{\bullet}$ is a filtration of $\mathcal{V}_{\mathcal{O}} := \mathcal{O}_S \otimes_{\mathbb{Q}} \mathcal{V}_{\mathbb{Q}}$ by holomorphic subbundles, and $Q\colon \mathcal{V}_{\mathbb{Q}} \times_S \mathcal{V}_{\mathbb{Q}} \to \mathbb{Q}(-n)_S$ is a flat bilinear form, such that Griffiths transversality condition $\nabla \mathcal{F}^p \subseteq \Omega^1_S \otimes_{\mathcal{O}_S} \mathcal{F}^{p-1}$ holds, and such that $\mathcal{V}$ induces a polarized $\mathbb{Q}$-Hodge structure $(\mathcal{V}_{\mathbb{Q},s}, \mathcal{F}^{\bullet}_s, Q_s)$ of weight n on every fibre.

Let $\pi\colon \widetilde{S} \to S$ be a universal covering, and choose a trivialization $\pi^*\mathcal{V}_{\mathbb{Q}} \cong \widetilde{S} \times V$. For $s \in S$, let $\mathrm{MT}_s \subseteq \mathrm{GL}(\mathcal{V}_{\mathbb{Q},s})$ denote the Mumford-Tate group of its fibre. The choice of a point $\tilde{s} \in \widetilde{S}$ with $\pi(\tilde{s}) = s$ gives an identification $\mathcal{V}_{\mathbb{Q},s} \cong V$, whence an injective homomorphism $i_{\tilde{s}}\colon \mathrm{MT}_s \hookrightarrow \mathrm{GL}(V)$.

There exists a countable union $\Sigma \subsetneq S$ of proper analytic subspaces of S with the following properties:

(i) for $s \in S \setminus \Sigma$, the image $M := \mathrm{Im}(i_{\tilde{s}}) \subseteq \mathrm{GL}(V)$ does not depend on s, nor on the choice of $\tilde{s}$,

(ii) for all s and $\tilde{s}$ as above with $s \in \Sigma$, the image of $i_{\tilde{s}}$ is a proper subgroup of M.

We call $S \setminus \Sigma$ the "Hodge-generic" locus. The group M in (i) is called the generic Mumford-Tate group of $\mathcal{V}$. More intrinsically, for any $s \in S \setminus \Sigma$ we refer to MT_s as "the" generic Mumford-Tate group of $\mathcal{V}$.

If S is a nonsingular complex algebraic variety, then Σ is a countable union of algebraic subvarieties of S; this was shown in [10].

1.3 From now on we assume that S is a connected, nonsingular complex algebraic variety and that $\mathcal{V}$ admits a $\mathbb{Z}$-structure (i.e., $\mathcal{V}_{\mathbb{Q}} = \mathcal{V}_{\mathbb{Z}} \otimes \mathbb{Q}$, where $\mathcal{V}_{\mathbb{Z}}$ underlies a polarizable variation of $\mathbb{Z}$-Hodge structure). Choose a base-point $s \in S$ and a point $\tilde{s} \in \widetilde{S}$ with $\pi(\tilde{s}) = s$. The local system $\mathcal{V}_{\mathbb{Q}}$ underlying $\mathcal{V}$ then corresponds to a representation $\rho\colon \pi_1(S, s) \to \mathrm{GL}(V)$, called the monodromy representation. The algebraic monodromy group is defined as the smallest algebraic subgroup of $\mathrm{GL}(V)$ defined over $\mathbb{Q}$ which contains the image of ρ. We write $H_s = H_{\mathrm{mon},s}$ for its connected component of the identity, called the connected algebraic monodromy group. Given the trivialization of $\pi^*\mathcal{V}_{\mathbb{Q}}$, the group $H_{\mathrm{mon},s} \subseteq \mathrm{GL}(V)$ is independent of the choice of s and $\tilde{s}$.

1.4 Theorem. *Assumptions as in 1.3 and notations as above.*

(i) The group $H_{\mathrm{mon},s}$ is a normal subgroup of the derived group M^{der}.

(ii) Suppose there is a point $t \in S$ such that MT_t is abelian (hence a torus). Then $H_{\mathrm{mon},s} = M^{\mathrm{der}}$.

For the first statement we refer to Deligne's paper [15]. The second statement was proven in the more general context of variation of mixed Hodge structure by André in [2].

§2 Subvarieties of Hodge type

2.1 For the basic theory of Shimura varieties, we refer to Deligne's papers [14] and [17]. We follow some notations and conventions of [17, Sect. 0], in particular we write a superscript ^{0} for algebraic connected components, a superscript $^+$ for analytic connected components, and if G is a reductive group over $\mathbb{Q}$ then we write $G(\mathbb{Q})_+$ for the intersection of $G(\mathbb{Q})$ and the inverse image of $G^{\mathrm{ad}}(\mathbb{R})^+$ under the adjoint map. If G is a reductive group then we write $Z(G)$ for its center.

A Shimura datum is a pair $(G, \mathfrak{X})$ consisting of an algebraic group G defined over $\mathbb{Q}$ and a $G(\mathbb{R})$-conjugacy class $\mathfrak{X} \subseteq \mathrm{Hom}(\mathbb{S}, G_{\mathbb{R}})$, satisfying the axioms [17, (2.1.1.1-3)]. We write $Sh_K(G, \mathfrak{X})$ for the Shimura variety (over $\mathbb{C}$) associated to a Shimura datum $(G, \mathfrak{X})$ and a compact open subgroup $K \subset G(\mathbb{A}_f)$; by definition we thus have

$$Sh_K(G, \mathfrak{X})(\mathbb{C}) = G(\mathbb{Q})\backslash \mathfrak{X} \times G(\mathbb{A}_f)/K\,.$$

We adopt the notational convention that symbols $\mathfrak{X}$, $\mathfrak{Y}$ etc. represent the conjugacy classes which are part of a Shimura datum, and that symbols X, Y etc. represent connected components (which are hermitian symmetric domains).

A morphism $f\colon (G_1, \mathfrak{X}_1) \to (G_2, \mathfrak{X}_2)$ of Shimura data is defined as a homomorphism $f\colon G_1 \to G_2$ of algebraic groups over $\mathbb{Q}$ which induces a map from $\mathfrak{X}_1$ to $\mathfrak{X}_2$. We call f a closed immersion if it identifies G_1 with a closed subgroup of G_2. If $f\colon (G_1, \mathfrak{X}_1) \to (G_2, \mathfrak{X}_2)$ is a morphism of Shimura data, and $K_1 \subseteq G_1(\mathbb{A}_f)$, $K_2 \subseteq G_2(\mathbb{A}_f)$ are compact open subgroups with $f(K_1) \subseteq K_2$, then we write

$$f_{(K_1,K_2)}\colon Sh_{K_1}(G_1, \mathfrak{X}_1) \to Sh_{K_2}(G_2, \mathfrak{X}_2)$$

for the morphism induced by f. In the particular case that $(G_1, \mathfrak{X}_1) = (G_2, \mathfrak{X}_2)$ and f is the identity, we write $Sh_{(K_1,K_2)}$ instead of $f_{(K_1,K_2)}$.

If $(G, \mathfrak{X})$ is a Shimura datum then we write $\mathfrak{X}^{\mathrm{ad}}$ for the $G^{\mathrm{ad}}(\mathbb{R})$-conjugacy class of homomorphisms containing the image of $\mathfrak{X}$ in $\mathrm{Hom}(\mathbb{S}, G^{\mathrm{ad}}_{\mathbb{R}})$. The pair $(G^{\mathrm{ad}}, \mathfrak{X}^{\mathrm{ad}})$ is again a Shimura datum, called the adjoint datum of $(G, \mathfrak{X})$. The natural map $\mathfrak{X} \to \mathfrak{X}^{\mathrm{ad}}$ identifies $\mathfrak{X}$ with the union of a number of components of $\mathfrak{X}^{\mathrm{ad}}$.

A point $x \in \mathfrak{X}$ is called a special point if there exists a torus $T \subset G$, defined over $\mathbb{Q}$, such that $h_x\colon \mathbb{S} \to G_{\mathbb{R}}$ factors through $T_{\mathbb{R}}$. A point of $Sh_K(G, \mathfrak{X})$ is called a special point if it is of the form $[x, \eta K]$ for some special point $x \in \mathfrak{X}$.

2.2 Let $K_1, K_2 \subset G(\mathbb{A}_f)$ be compact open subgroups, let $g \in G(\mathbb{A}_f)$ and write $K' = K_1 \cap gK_2g^{-1}$. The Hecke correspondence $\mathcal{T}_g$ from $Sh_{K_1}(G, \mathfrak{X})$ to $Sh_{K_2}(G, \mathfrak{X})$ is defined by the diagram

$$\begin{array}{ccc} Sh_{K'}(G,\mathfrak{X}) & \xrightarrow{\pi_2} & Sh_{K_2}(G,\mathfrak{X}), \qquad \pi_2\colon [x,\theta K'] \mapsto [x,\theta g K_2] \\ {\scriptstyle \pi_1 := Sh_{(K',K_1)}} \downarrow & & \\ Sh_{K_1}(G,\mathfrak{X}) & & \end{array}$$

In general, we will not indicate K_1 and K_2 in the notation; this should not cause any confusion. Even though the $\mathcal{T}_g$ are correspondences, we will apply the usual terminology for morphisms to them. In particular, for a subvariety Z of $Sh_{K_1}(G, \mathfrak{X})$ we write $\mathcal{T}_g(Z) := \pi_2(\pi_1^{-1}(Z)) \subseteq Sh_{K_2}(G, \mathfrak{X})$ for the image of Z in the sense of correspondences.

2.3 Let $(G, \mathfrak{X})$ be a Shimura datum, and consider a representation $\xi\colon G \to \mathrm{GL}(V)$ such that the weight $w\colon \mathbb{G}_{\mathrm{m},\mathbb{C}} \to (G/\mathrm{Ker}(\xi))_{\mathbb{C}}$ is defined over $\mathbb{Q}$. Then ξ gives rise to a polarizable VHS over $\mathfrak{X}$ with underlying bundle $\mathfrak{X} \times V$. The axioms of a Shimura datum imply that the generic Mumford-Tate group of this VHS is a normal subgroup of $\xi(G)$ containing $\xi(G)^{\mathrm{der}}$.

Let C be the center of $G/\mathrm{Ker}(\xi)$, and assume that C^0 is an almost direct product $C^0 = C_1 \cdot C_2$, where C_1 is a $\mathbb{Q}$-split torus and $C_2(\mathbb{R})$ is compact. Using arguments as in [22, Prop. II.3.3(a)], one can show that for K sufficiently small, the VHS over $\mathfrak{X}$ descends to a polarizable VHS $\mathcal{V}(\xi)$ on $Sh_K(G, \mathfrak{X})$. Notice that both the condition on the weight and the condition on C^0 are satisfied if ξ factors through the adjoint group G^{ad}.

2.4 Let $H \subseteq G$ be an algebraic subgroup defined over $\mathbb{Q}$, and let

$$\mathfrak{Y}_H = \{x \in \mathfrak{X} \mid h_x\colon \mathbb{S} \to G_{\mathbb{R}} \text{ factors through } H_{\mathbb{R}}\}\,.$$

By using the fact that there are finitely many $H(\mathbb{R})$-conjugacy classes of maximal tori $T \subseteq H_{\mathbb{R}}$ defined over $\mathbb{R}$, and an argument similar to the proof of [17, Lemma 1.2.4], one can show that $\mathfrak{Y}_H$ is a finite union of $H(\mathbb{R})$-conjugacy classes.

2.5 Definition. An irreducible algebraic subvariety $S \subseteq Sh_K(G,\mathfrak{X})_{\mathbb{C}}$ is called a subvariety of Hodge type if there exist an algebraic subgroup $H \subseteq G$ (defined over $\mathbb{Q}$), an element $\eta \in G(\mathbb{A}_f)$ and a connected component Y_H of $\mathfrak{Y}_H$ such that $S(\mathbb{C})$ is the image of $Y_H \times \eta K$ in $Sh_K(G,X)(\mathbb{C}) = G(\mathbb{Q})\backslash\mathfrak{X} \times G(\mathbb{A}_f)/K$.

Observe that if $K_1 \subseteq K_2$ are two compact open subgroups of $G(\mathbb{A}_f)$, then $S \subseteq Sh_{K_2}(G,\mathfrak{X})$ is a subvariety of Hodge type if and only if it is the image under the map $Sh_{(K_1,K_2)}$ of a subvariety of Hodge type of $Sh_{K_1}(G,\mathfrak{X})$.

2.6 Remark. Let $f\colon (G_1,\mathfrak{X}_1) \to (G_2,\mathfrak{X}_2)$ be a closed immersion of Shimura data, and let $K_1 \subset G_1(\mathbb{A}_f)$ and $K_2 \subset G_2(\mathbb{A}_f)$ be compact open subgroups with $f(K_1) \subseteq K_2$. If $S \subseteq Sh_{K_2}(G_2,\mathfrak{X}_2)$ is an irreducible component of the image of the associated map $f_{(K_1,K_2)}\colon Sh_{K_1}(G_1,\mathfrak{X}_1) \to Sh_{K_2}(G_2,\mathfrak{X}_2)$, then S is a subvariety of Hodge type. Let us call an S obtained in this way a subvariety of Shimura type.

In general it is not true that all subvarieties of Hodge type are of Shimura type. The point is that S may lie in a component of $Sh_{K_2}(G_2,\mathfrak{X}_2)$ which is not in the image of $f_{(K_1,K_2)}$. Up to a Hecke correspondence, all subvarieties of Hodge type are of Shimura type, however. More precisely: let S be the image of $Y_H \times \eta K$ as in Def. 2.5, and write $\mathfrak{C} \subset \mathfrak{Y}_H$ for the $H(\mathbb{R})$-conjugacy class containing Y_H. Then the pair $(H,\mathfrak{C})$ is a Shimura datum, and if $K' \subset H(\mathbb{A}_f)$ is a compact open subgroup contained in K, the inclusion $H \subset G$ induces a morphism $Sh_{K'}(H,\mathfrak{C}) \to Sh_K(G,\mathfrak{X})$. Writing S^e for the image of $Y_H \times K$ in $Sh_K(G,X)$, we see that S^e is of Shimura type, and that S is an irreducible component of $\mathcal{T}_\eta(S^e)$. In particular, this shows that for every algebraic subgroup $H \subset G$ over $\mathbb{Q}$ and every $\eta \in G(\mathbb{A}_f)$, the image of $\mathfrak{Y}_H \times \eta K$ in $Sh_K(G,X)$ is an algebraic subvariety.

2.7 Lemma. *Let H be a subgroup of G such that $\mathfrak{Y}_H$ is non-empty. Then $\mathfrak{Y}_H = \mathfrak{Y}_{Z(G)\cdot H}$.*

Proof. Since h_x factors through $Z(G)\cdot H$ if and only if it factors through the connected component $(Z(G)\cdot H)^0 \subseteq Z(G)^0 \cdot H$, it suffices to show that $\mathfrak{Y}_H = \mathfrak{Y}_{Z(G)^0\cdot H}$. Write

$$Z^0 = Z(G)^0\,, \quad T = (Z^0 \cap H)\cdot(Z^0 \cap G^{\mathrm{der}}) \subseteq Z^0\,,$$

and, for $y \in Y_{Z(G)^0\cdot H}$, consider the homomorphisms

$$f_1\colon\ \mathbb{S} \xrightarrow{h_y} G_{\mathbb{R}} \longrightarrow G^{\mathrm{ab}}_{\mathbb{R}} \cong Z^0/(Z^0 \cap G^{\mathrm{der}}) \longrightarrow Z^0/T\,,$$

$$f_2\colon\ \mathbb{S} \xrightarrow{h_y} Z^0\cdot H \longrightarrow (Z^0\cdot H)/H \cong Z^0/(Z^0\cap H), \quad \text{and}$$

$$\pi\colon\ Z^0/(Z^0\cap H) \longrightarrow Z^0/T\,.$$

One easily checks that $\pi \circ f_2 = f_1$. The image of f_1 does not depend on y, since the image of $\mathbb{S}$ in $G^{\mathrm{ab}}_{\mathbb{R}}$ is already independent of y. The condition that $\mathfrak{Y}_H$ is non-empty means that there exists a $y \in \mathfrak{Y}_{H\cdot Z(G)^0}$ for which the image of f_2 is the identity element in $Z^0/(Z^0\cap H)$. Therefore the image of f_2 is contained in $\mathrm{Ker}(\pi)$ for every $y \in \mathfrak{Y}_{H\cdot Z(G)^0}$. Since π is an isogeny ($Z^0 \cap G^{\mathrm{der}}$ being a finite group) and $\mathbb{S}$ is connected, we conclude that f_2 is trivial for every y, which proves the lemma. ☐

The condition that $\mathfrak{Y}_H$ is non-empty puts strong restrictions on H; it implies, for instance, that H^0 is a reductive subgroup of G.

2.8 Proposition. *Given a subvariety $S \subseteq Sh_K(G, \mathfrak{X})$ of Hodge type, there exists a compact open subgroup K' of $G(\mathbb{A}_f)$ contained in K, a representation $\xi\colon G \to \mathrm{GL}(V)$ which is induced from a faithful representation of G^{ad}, and an algebraic subgroup $M \subseteq \mathrm{GL}(V)$, such that*

1. *ξ induces a polarizable VHS $\mathcal{V}(\xi)$ over $Sh_{K'}(G, \mathfrak{X})$,*
2. *S is the image under the map $Sh_{(K',K)}$ of an irreducible subvariety $S' \subseteq Sh_{K'}(G, \mathfrak{X})$ such that S' is a maximal irreducible subvariety with generic Mumford-Tate group M.*

Note that it makes sense to state that $\mathcal{V}(\xi)|_{S'}$ has generic Mumford-Tate group M, since the irreducible component of $Sh_{K'}(G, \mathfrak{X})$ containing S' is a quotient of X, and over X the bundle underlying $\mathcal{V}(\xi)$ is just $\mathfrak{X} \times V$.

Proof. Using the lemma we can describe S as the image of some $Y_H \times \eta K$, where H contains $Z(G)$. Choose a faithful representation $\xi^{\mathrm{ad}}\colon G^{\mathrm{ad}} \to \mathrm{GL}(V)$, and write ξ for the induced representation of G. For $K' \subseteq K$ sufficiently small we get a polarizable VHS $\mathcal{V}(\xi)$ over $Sh_{K'}(G, \mathfrak{X})$, as explained in 2.3. Let S' be an irreducible component of $Sh^{-1}_{(K',K)}(S)$, and define $M \subseteq \xi(H)$ as the generic Mumford-Tate group of $\mathcal{V}(\xi)|_{S'}$. Clearly, $\xi^{-1}(M)^0 \subseteq H$ and $Y_H \subseteq \mathfrak{Y}_{\xi^{-1}(M)} \subseteq \mathfrak{Y}_H$, hence Y_H is a connected component of $\mathfrak{Y}_{\xi^{-1}(M)}$. The proposition readily follows. □

2.9 Let Z be an irreducible algebraic subvariety of a Shimura variety $Sh_K(G, \mathfrak{X})$. There exists a unique smallest subvariety of Hodge type, say S, containing Z. (Notice that the intersection of two subvarieties of Hodge type is again of Hodge type.) By definition, S is an irreducible component of the image of $\mathfrak{Y}_M \times \eta K$ in $Sh_K(G, \mathfrak{X})$, where $M \subseteq G$ is an algebraic subgroup (over $\mathbb{Q}$) and $\eta \in G(\mathbb{A}_f)$. If $Y_M \subset \mathfrak{Y}_M$ is a connected component such that S is the image of $Y_M \times \eta K$ then we write $S = S_{\eta K}(Y_M)$.

This description does not uniquely determine the group M. However, as we have seen, we can take for M the "generic Mumford-Tate group on Z". More precisely, let $K' \subset G(\mathbb{A}_f)$ be a compact open subgroup contained in K and let $\xi\colon G \to \mathrm{GL}(V)$ be a representation such that we obtain a polarizable VHS $\mathcal{V}(\xi)$ over $Sh_{K'}(G, \mathfrak{X})$. Let $Z' \hookrightarrow S'$ be irreducible components of $Sh^{-1}_{(K',K)}(Z)$ and $Sh^{-1}_{(K',K)}(S)$ respectively. The generic Mumford-Tate group MT of $\mathcal{V}(\xi)|_{Z'}$ is equal to that of $\mathcal{V}(\xi)|_{S'}$ and we may choose M in the above such that MT is conjugated to $\xi(M)$ (for all representations ξ which induce a VHS for K' sufficiently small). Up to conjugation by elements of $G(\mathbb{Q})$ this uniquely determines $Z(G) \cdot M \subseteq G$.

§3 Decomposition of the adjoint group

3.1 Consider a closed immersion $i\colon (M, \mathfrak{Y}) \hookrightarrow (G, \mathfrak{X})$ of Shimura data. Let $M^{\mathrm{ad}} = M_1 \times M_2$ be a decomposition of the adjoint group of M. (We do not assume M_1 and M_2 to be non-trivial or $\mathbb{Q}$-simple.) There is a corresponding decomposition $\mathfrak{Y}^{\mathrm{ad}} = \mathfrak{Y}_1 \times \mathfrak{Y}_2$, where $\mathfrak{Y}_i$ $(i \in \{1, 2\})$ is a union of Hermitian symmetric domains, and $M_i(\mathbb{R})$ acts transitively on $\mathfrak{Y}_i$. One easily checks that $(M_i, \mathfrak{Y}_i)$ is a Shimura datum, so we have a decomposition of Shimura data $(M^{\mathrm{ad}}, \mathfrak{Y}^{\mathrm{ad}}) = (M_1, \mathfrak{Y}_1) \times (M_2, \mathfrak{Y}_2)$.

Choose compact open subgroups $C_i \subset M_i(\mathbb{A}_f)$, and $C \subset M(\mathbb{A}_f)$, with $\mathrm{ad}(C) \subseteq C_1 \times C_2$. For C_1 and C_2 sufficiently small the associated morphism

$$\mathrm{ad}_{(C, C_1 \times C_2)}\colon \mathit{Sh}_C(M, \mathfrak{Y}) \longrightarrow \mathit{Sh}_{C_1 \times C_2}(M^{\mathrm{ad}}, \mathfrak{Y}^{\mathrm{ad}}) = \mathit{Sh}_{C_1}(M_1, \mathfrak{Y}_1) \times \mathit{Sh}_{C_2}(M_2, \mathfrak{Y}_2)$$

is finite étale on irreducible components. Given a connected component $Y_1 \subseteq \mathfrak{Y}_1$, a point $y_2 \in \mathfrak{Y}_2$ and a class $\theta C \in M(\mathbb{A}_f)/C$, let

$$S_{\theta C}(Y_1, y_2) \subseteq \mathit{Sh}_C(M, \mathfrak{Y})$$

denote the image of $(Y_1 \times \{y_2\}) \times \theta C$ in $\mathit{Sh}_C(M, \mathfrak{Y})$.[1] If $\theta C_1 \times \theta C_2$ is the image of θC in $M_1(\mathbb{A}_f)/C_1 \times M_2(\mathbb{A}_f)/C_2$, then $S_{\theta C}(Y_1, y_2)$ is an irreducible component of the inverse image of $\mathit{Sh}_{C_1}(M_1, \mathfrak{Y}_1) \times [y_2, \theta_2 C_2]$ under $\mathrm{ad}_{(C, C_1 \times C_2)}$. In other words, the Shimura variety $\mathit{Sh}_{C_1 \times C_2}(M^{\mathrm{ad}}, \mathfrak{Y}^{\mathrm{ad}})$ is a disjoint union

$$\mathit{Sh}_{C_1 \times C_2}(M^{\mathrm{ad}}, \mathfrak{Y}^{\mathrm{ad}}) = \mathit{Sh}_{C_1}(M_1, \mathfrak{Y}_1) \times \mathit{Sh}_{C_2}(M_2, \mathfrak{Y}_2) = \coprod_{i \in I, j \in J} (\Gamma_i \backslash Y_1) \times (\Gamma_j \backslash Y_2)$$

of product varieties, and $S_{\theta C}(Y_1, y_2)$ is an irreducible subvariety of $\mathit{Sh}_C(M, \mathfrak{Y})$ covering some $(\Gamma_i \backslash Y_1) \times [y_2]$.

More generally, if $K \subset G(\mathbb{A}_f)$ is a compact open subgroup and $\eta K \in G(\mathbb{A}_f)/K$, then we define

$$S_{\eta K}(Y_1, y_2)$$

as the image of $(Y_1 \times \{y_2\}) \times \eta K$ in $\mathit{Sh}_K(G, \mathfrak{X})$. Notice that $S_{\eta K}(Y_1, y_2)$ is an algebraic subvariety. This follows from the remark that $S_{\eta K}(Y_1, y_2)$ is an irreducible component of $\mathcal{T}_\eta(S_{eK}(Y_1, y_2))$, and for C small enough $S_{eK}(Y_1, y_2)$ is the image of $S_{eC}(Y_1, y_2) \subseteq \mathit{Sh}_C(M, \mathfrak{Y})$ under the finite morphism $i_{(C,K)}$.

The subvarieties of the form $S_{\eta K}(Y_1, y_2) \subseteq \mathit{Sh}_K(G, \mathfrak{X})$ are totally geodesic (see Section 4.1), since $Y_1 \times \{y_2\}$ is a complete totally geodesic submanifold of $\mathfrak{Y} = \mathfrak{Y}_1 \times \mathfrak{Y}_2$, and $\mathfrak{Y}$ is totally geodesic in $\mathfrak{X}$. Note that $S_{\eta K}(Y_1, y_2)$ contains special points if and only if y_2 is a special point of $\mathfrak{Y}_2$, in which case the special points are dense.

3.2 As before, let $(G, \mathfrak{X})$ be a Shimura datum and let $K \subset G(\mathbb{A}_f)$ be a compact open subgroup. We consider an irreducible algebraic subvariety Z of $\mathit{Sh}_K(G, \mathfrak{X})$. As discussed in Section 2.9, there is a smallest subvariety of Hodge type containing Z, which we denote by $S_{\eta K}(Y_M)$. It corresponds to a closed immersion $i\colon (M, \mathfrak{Y}_M) \hookrightarrow (G, \mathfrak{X})$ of Shimura data, a connected component $Y_M \subseteq \mathfrak{Y}_M$ and a class $\eta K \in G(\mathbb{A}_f)/K$. If there is no risk of confusion we simply write $Y = Y_M$ and $S = S_{\eta K}(Y_M)$. We write $j\colon Z \hookrightarrow S$ for the inclusion map.

Possibly after replacing K by a subgroup K' of finite index and Z by an irreducible component of its preimage in $S_{\eta K'}(Y_M)$ we may assume that the following conditions hold.

[1] When using this notation, we shall always tacitly assume that $Y_1 \times \{y_2\} \subseteq \mathfrak{Y} \subseteq \mathfrak{Y}^{\mathrm{ad}}$. (Without this assumption the definition obviously would not make any sense.)

1. there exists a representation $\xi\colon G \to \mathrm{GL}(V)$ with $\mathrm{Ker}(\xi) \subseteq Z(G)$ which induces a polarizable VHS $\mathcal{V}(\xi)$ on $Sh_K(G, \mathfrak{X})$ such that S is a maximal irreducible subvariety with generic Mumford-Tate group $\xi(M)$ (cf. Proposition 2.8).

2. K is neat; in particular, $Sh_K(G, \mathfrak{X})$ is a union of quotients $\Gamma_i \backslash X$ such that the natural maps $X \twoheadrightarrow \Gamma_i \backslash X$ are topological coverings and the algebraic monodromy group associated to the VHS $\mathcal{V}(\xi)$ over $\Gamma_i \backslash X$ is connected.

3. the natural map $u_S\colon Y \twoheadrightarrow S = S_{\eta K}(Y)$ is a topological covering.

For the last condition we need a lemma.

3.3 Lemma. *(i) For K sufficiently small the natural map $u_S\colon Y = Y_M \twoheadrightarrow S = S_{\eta K}(Y_M)$ is a topological covering.*

(ii) Let a decomposition $(M^{\mathrm{ad}}, \mathfrak{Y}^{\mathrm{ad}}) = (M_1, \mathfrak{Y}_1) \times (M_2, \mathfrak{Y}_2)$ be given. For K sufficiently small the map $Y_1 \twoheadrightarrow S_{\eta K}(Y_1, y_2)$ is a covering map for every $y_2 \in \mathfrak{Y}_2$ and $\eta \in G(\mathbb{A}_f)$.

Proof. Take a compact open subgroup $C \subseteq M(\mathbb{A}_f)$ such that $Sh_C(M, \mathfrak{Y})$ is non-singular. By [14, Prop. 1.15] there exists a compact open subgroup $K \subset G(\mathbb{A}_f)$ such that $i(C) \subseteq K$ and such that $i_{(C,K)}\colon Sh_C(M, \mathfrak{Y}) \to Sh_K(G, \mathfrak{X})$ is a closed immersion. For this choice of C and K the map $Y \twoheadrightarrow S_{eK}(Y)$ therefore is a topological covering.

Let $K' = K \cap \eta^{-1} K \eta$. Clearly, both K' and $\eta K' \eta^{-1}$ are contained in K; in particular, $Y \twoheadrightarrow S_{e(\eta K' \eta^{-1})}(Y)$ is a covering. Let X be the component of $\mathfrak{X}$ containing Y. As one easily verifies, the map $S_{e(\eta K' \eta^{-1})}(X) \to S_{\eta K'}(X)$ obtained by sending the class $[x, eK]$ to the class $[x, \eta K]$ is an isomorphism, compatible with the uniformization maps from X. Restricting this to Y we conclude that $Y \twoheadrightarrow S_{\eta K'}(Y)$ is a covering, which proves the first part of the lemma.

The second statement easily follows from the first one and [6, Cor. 8.10]. □

3.4 Let $\widetilde{Z}$ be a connected component of $u_S^{-1}(Z)$. The map $u_Z = u_S|_{\widetilde{Z}}\colon \widetilde{Z} \to Z$ is again a topological covering. If $\mathrm{Cov}(u_S)$ and $\mathrm{Cov}(u_Z)$ denote the groups of covering transformations, then $\mathrm{Cov}(u_Z) = \{\gamma \in \mathrm{Cov}(u_S) \mid \gamma \widetilde{Z} = \widetilde{Z}\}$. In general the analytic space $\widetilde{Z}$ is not irreducible[2], and in some arguments to follow this causes problems. To circumvent these, we consider the normalization $n\colon Z^n \to Z$ of Z. Let $u_{Z^n}\colon \widetilde{Z^n} \twoheadrightarrow Z^n$ be a universal covering of Z^n. (Caution: our notations may be somewhat misleading, since $\widetilde{Z}$ need not be a *universal* covering of Z, whereas we do write $\widetilde{Z^n}$ for a universal covering of Z^n.) The analytic space $\widetilde{Z^n}$ is connected and normal; in particular it is irreducible.

Let $\mathcal{C} \subseteq \widetilde{Z}$ be an irreducible component. Choose a Hodge generic and regular base point $z \in Z$, i.e., a regular point outside the locus Σ as in Section 1.2 (applied to the VHS $\mathcal{V}(\xi)$ over S). This is possible, since $Z \subseteq \Sigma$ would contradict the fact that S is the smallest subvariety of Hodge type containing Z. Also choose base points $\tilde{z} \in \mathcal{C}$, $\zeta \in Z^n$ and $\tilde{\zeta} \in \widetilde{Z^n}$ with $u_Z(\tilde{z}) = z$, $n(\zeta) = z$ and $u_{Z^n}(\tilde{\zeta}) = \zeta$. There is a well-determined morphism $\tilde{n}\colon \widetilde{Z^n} \to \widetilde{Z}$ with $\tilde{n}(\tilde{\zeta}) = \tilde{z}$ and $u_Z \circ \tilde{n} = n \circ u_{Z^n}$. We have $\tilde{n}(\widetilde{Z^n}) = \mathcal{C} \subseteq \widetilde{Z}$.

[2] We do not know whether it is possible in general to choose K small enough such that $\widetilde{Z}$ is irreducible.

In a diagram, the situation looks as follows:

$$\begin{array}{ccccccccc} \widetilde{Z^n} & \xrightarrow{\tilde{n}} & \mathcal{C} & \hookrightarrow & \widetilde{Z} & \hookrightarrow & Y & \hookrightarrow & \mathfrak{X} \\ {\scriptstyle u_{Z^n}}\downarrow & & & \searrow & \downarrow{\scriptstyle u_Z} & & \downarrow{\scriptstyle u_S} & & \downarrow \\ Z^n & & \xrightarrow{n} & & Z & \hookrightarrow & S = S_{\eta K}(Y) & \hookrightarrow & Sh_K(G,\mathfrak{X}) \end{array}$$

The choice of the point $\tilde{z}$ above z gives an identification of the fibre $\mathcal{V}(\xi)_z$ with V, and we identify the Mumford-Tate group at z with $\xi(M) \subseteq \mathrm{GL}(V)$. This is also the generic Mumford-Tate group of the VHS $n^*\mathcal{V}(\xi)$ over Z^n, via the identification $(n^*\mathcal{V}(\xi))_\zeta \cong \mathcal{V}(\xi)_z = V$. Since $\mathrm{Ker}(\xi) \subseteq Z(G)$ there is a natural surjective homomorphism $f\colon \xi(M) \to M^{\mathrm{ad}}$, and composing the monodromy representation $\rho_S\colon \pi_1(S,z) \to \xi(M)(\mathbb{Q}) \subseteq \mathrm{GL}(V)$ with f we obtain a homomorphism $f \circ \rho_S\colon \pi_1(S,z) \to M^{\mathrm{ad}}(\mathbb{Q}) \subset M^{\mathrm{ad}}(\mathbb{R})$.

3.5 Lemma. *We have* $\mathrm{Im}(f \circ \rho_S) \subseteq M^{\mathrm{ad}}(\mathbb{R})^+$, *and there is a commutative diagram*

$$\begin{array}{ccccccccc} \pi_1(Z^n,\zeta) & & \xrightarrow{n_*} & & \pi_1(Z,z) & \xrightarrow{j_*} & \pi_1(S,z) & \xrightarrow{f\circ\rho} & M^{\mathrm{ad}}(\mathbb{R})^+ \\ \downarrow{\scriptstyle\cong} & & & & \downarrow & & \downarrow{\scriptstyle\cong} & & \downarrow \\ \mathrm{Cov}(u_{Z^n}) & \rightarrow & \mathrm{Cov}(u_Z)_{\mathcal{C}} & \hookrightarrow & \mathrm{Cov}(u_Z) & \hookrightarrow & \mathrm{Cov}(u_S) & \hookrightarrow & \mathrm{Aut}(Y) \end{array}$$

where $\mathrm{Cov}(u_Z)_{\mathcal{C}} = \{\gamma \in \mathrm{Cov}(u_Z) \mid \gamma\mathcal{C} = \mathcal{C}\}$.

Proof. It is clear that $\pi_1(Z^n,\zeta) \cong \mathrm{Cov}(u_{Z^n})$ maps into $\mathrm{Cov}(u_Z)_{\mathcal{C}}$ and that the two diagrams on the left are commutative. We therefore only have to consider the right-hand square. Let $\Gamma\backslash X$ be the irreducible component of $Sh_K(G,\mathfrak{X})$ containing Z, where Γ is of the form $\Gamma = G(\mathbb{Q})_+ \cap gKg^{-1}$. The local system $\mathcal{V}(\xi)$ over $\Gamma\backslash X$ is the quotient of the trivial bundle $X \times V$ over X under the action of Γ given by $\gamma(x,v) = (\gamma \cdot x, \xi(\gamma)\cdot v)$.

Take $\alpha \in \pi_1(S,z)$ and let $\gamma \in \Gamma$ be an element mapping to $i_*(\alpha) \in \pi_1(\Gamma\backslash X, z)$. Then $\gamma Y = Y$, and the image of α in $\mathrm{Aut}(Y)$ is given by the action of γ. The given description of $\mathcal{V}(\xi)$ shows that $\rho_S(\alpha) = \xi(\gamma)$. In particular, $\gamma \in Z(G)\cdot M$. It now readily follows that $f \circ \rho_S(\alpha) \in M^{\mathrm{ad}}(\mathbb{R})^+$ (since the action of γ stabilizes Y), and that the right-hand square is commutative. □

3.6 As in the proof of the lemma, let $\Gamma\backslash X$ be the irreducible component of $Sh_K(G,\mathfrak{X})$ containing Z. The choice of a Γ-stable lattice in V induces a $\mathbb{Z}$-structure on $\mathcal{V}(\xi)$, which enables us to apply Theorem 1.4. The connected algebraic monodromy group $H_\zeta = H_{\mathrm{mon},\zeta}$ associated to the VHS $n^*(\mathcal{V}(\xi)_{|Z})$ is therefore a normal subgroup of $\xi(M)^{\mathrm{der}} = \xi(M^{\mathrm{der}})$. Since M is reductive, we can find a normal algebraic subgroup $H_2 \triangleleft M$ (defined over $\mathbb{Q}$) such that M is the almost direct product of $\xi^{-1}(H_\zeta)$ and H_2. In this way we obtain a decomposition

$$(M^{\mathrm{ad}}, \mathfrak{Y}^{\mathrm{ad}}) = (H_\zeta^{\mathrm{ad}}, \mathfrak{Y}_1) \times (H_2^{\mathrm{ad}}, \mathfrak{Y}_2)\,.$$

3.7 Proposition. *The image of* $\mathcal{C}$ *under the projection map* $\mathrm{pr}_2\colon \mathfrak{Y} \to \mathfrak{Y}_2$ *is a single point, say* $y_2 \in \mathfrak{Y}_2$. *We have* $Z \subseteq S_{\eta K}(Y_1, y_2)$ *for some connected component* $Y_1 \subseteq \mathfrak{Y}_1$ *and a class* $\eta K \in G(\mathbb{A}_f)/K$.

Proof. It follows from the lemma that $\mathrm{Cov}(u_{Z^n})$ acts trivially on $\mathfrak{Y}_2$, hence the composition $\widetilde{Z^n} \twoheadrightarrow \mathcal{C} \subseteq \mathfrak{Y} \twoheadrightarrow \mathfrak{Y}_2$ factors through Z^n. Because the components of $\mathfrak{Y}_2$ have a realization as a bounded domain in some $\mathbb{C}^N$, the map $Z^n \to \mathfrak{Y}_2$ is given by an N-tuple of bounded holomorphic functions. Since Z^n is a connected quasi-projective variety these must be constant functions hence the image of $\mathcal{C}$ is a single point.

The last assertion is an immediate consequence of the first. □

If Z contains a regular special point then $H_\zeta = \xi(M)^{\mathrm{der}}$, by the second statement of Theorem 1.4. This means that $\mathfrak{Y} \cong \mathfrak{Y}_1$, $\mathfrak{Y}_2$ is a point and that $S_{\eta K}(Y_1, y_2) = S$ is a subvariety of Hodge type. In this case, the proposition does not give us any information. However, the very fact that $H_\zeta = \xi(M)^{\mathrm{der}}$ can be used to establish a second decomposition of $(M^{\mathrm{ad}}, \mathfrak{Y}^{\mathrm{ad}})$.

Consider the group $\{m \in M(\mathbb{Q})_+ \mid m\mathcal{C} = \mathcal{C}\}$, and write $\mathcal{N}$ for its closure inside $M(\mathbb{R})$ for the analytic topology. By Cartan's theorem $\mathcal{N}$ and its connected component of the identity $\mathcal{N}^+$ are Lie subgroups of $M(\mathbb{R})$. Clearly they are contained in $M(\mathbb{R})_{\mathcal{C}} = \{m \in M(\mathbb{R}) \mid m\mathcal{C} = \mathcal{C}\}$.

3.8 Proposition. *Assume that Z contains a non-singular special point. Then there exists a normal, reductive algebraic subgroup $N_Z \triangleleft M$, defined over $\mathbb{Q}$, such that $\mathcal{N}^+ = N_Z(\mathbb{R})^+$.*

Proof. The center $Z(M)(\mathbb{R})$ of $M(\mathbb{R})$ acts trivially on $\mathfrak{Y}$, and $Z(M)(\mathbb{Q})$ is analytically dense in $Z(M)(\mathbb{R})$, so $Z(M)(\mathbb{R}) \subseteq \mathcal{N}$. Furthermore, Lemma 3.5 shows that

$$\mathrm{ad}^{-1}\Big(\mathrm{Im}(f \circ \rho_S \circ j_* \circ n_*\colon \pi_1(Z^n, \zeta) \to M^{\mathrm{ad}}(\mathbb{Q})^+)\Big) \subseteq \mathcal{N}\,,$$

and since $H_\zeta = \xi(M)^{\mathrm{der}}$ it follows that $\mathcal{N}$ is Zariski dense in $M_{\mathbb{R}}$.

Let $\varphi\colon M_{\mathbb{R}} \to \mathrm{GL}(W)$ be a finite-dimensional irreducible representation of $M_{\mathbb{R}}$ with $\mathrm{Ker}(\varphi) \subseteq Z(M)$ (which exists, since M is reductive). Let $W' \subseteq W$ be the largest fully reducible $\mathcal{N}^+$-submodule of W. Then W' is an $\mathcal{N}$-submodule, since $\mathcal{N}$ normalizes $\mathcal{N}^+$. But $\mathcal{N}$ is Zariski dense in $M_{\mathbb{R}}$, so $W' = W$. Therefore, $\mathcal{N}^+/(\mathcal{N}^+ \cap \mathrm{Ker}(\varphi))$ has a faithful, fully reducible representation. Since $\mathcal{N}^+ \cap \mathrm{Ker}(\varphi)$ is contained in the center of $\mathcal{N}^+$, this implies that $\mathcal{N}^+$ is analytically reductive, i.e., $\mathrm{Lie}(\mathcal{N}^+)$ is reductive.

Write $\mathfrak{n} = \mathrm{Lie}(\mathcal{N}^+)$, which can be decomposed as $\mathfrak{n} = \mathfrak{c} \oplus \mathfrak{n}^{\mathrm{der}}$, where $\mathfrak{c}$ is the center and $\mathfrak{n}^{\mathrm{der}}$ is the derived algebra. From [11, Chap. II, Thm. 15] (alternatively, [7, Chap. II, Cor. 7.9]) we know that $\mathfrak{n}^{\mathrm{der}}$ is algebraic, so $\mathfrak{n}^{\mathrm{alg}} = \mathfrak{c}^{\mathrm{alg}} \oplus \mathfrak{n}^{\mathrm{der}}$.

Let $\mathcal{N}^{+,\mathrm{alg}} \subseteq M_{\mathbb{R}}$ be the algebraic envelope of $\mathcal{N}^+$, and let $\mathfrak{N}$ be the normalizer of $\mathcal{N}^{+,\mathrm{alg}}$ inside $M_{\mathbb{R}}$. Clearly, $\mathcal{N} \subseteq \mathfrak{N}(\mathbb{R})$. On the other hand, $\mathfrak{N}$ is an algebraic subgroup of $M_{\mathbb{R}}$ and $\mathcal{N}$ is Zariski dense, so $\mathfrak{N} = M_{\mathbb{R}}$ and $\mathcal{N}^{+,\mathrm{alg}}$ is a normal subgroup. This implies that $\mathfrak{n}^{\mathrm{alg}} = \mathrm{Lie}(\mathcal{N}^{+,\mathrm{alg}})$ is an ideal of $\mathrm{Lie}(M_{\mathbb{R}})$, hence $\mathfrak{c} \subseteq \mathfrak{c}^{\mathrm{alg}} \subseteq \mathrm{Lie}(Z(M)(\mathbb{R}))$. But, as remarked above, $Z(M)(\mathbb{R}) \subseteq \mathcal{N}$, so $\mathfrak{c} = \mathfrak{c}^{\mathrm{alg}} = \mathrm{Lie}(Z(M)(\mathbb{R}))$. We conclude that $\mathcal{N}^{+,\mathrm{alg}}$ is a normal, reductive algebraic subgroup of $M_{\mathbb{R}}$ and $\mathcal{N}^+$ is the connected component of the identity of the Lie group $\mathcal{N}^{+,\mathrm{alg}}(\mathbb{R})$.

Since $\mathcal{N}^+$ is open in $\mathcal{N}$ and $\{m \in M(\mathbb{Q})_+ \mid m\mathcal{C} = \mathcal{C}\}$ is analytically dense in $\mathcal{N}$, the group $\{m \in M(\mathbb{Q})_+ \cap \mathcal{N}^+ \mid m\mathcal{C} = \mathcal{C}\}$ is dense in $\mathcal{N}^+$ for the analytic topology and hence is Zariski dense in $\mathcal{N}^{+,\mathrm{alg}}$. Since it consists of $\mathbb{Q}$-valued points of M we conclude that $\mathcal{N}^{+,\mathrm{alg}}$ is defined over $\mathbb{Q}$, which proves the proposition. □

In the next section we need the following, similar, statement.

3.9 Variant. *Consider the inclusion $Z \subseteq S_{\eta K}(Y_1, y_2)$ as in Proposition 3.7, corresponding to the decomposition $(M^{\mathrm{ad}}, \mathfrak{Y}^{\mathrm{ad}}) = (H_\zeta^{\mathrm{ad}}, \mathfrak{Y}_1) \times (H_2^{\mathrm{ad}}, \mathfrak{Y}_2)$ and let $\mathcal{C} \subseteq \widetilde{Z}$ be an irreducible (analytic) component, as introduced after Lemma 3.3. There exists a normal algebraic subgroup $H_{\zeta,\mathcal{C}} \lhd H_{\zeta,\mathbb{R}}$ such that $H_{\zeta,\mathcal{C}}(\mathbb{R})^+ = \{h \in H_\zeta(\mathbb{R}) \mid h\mathcal{C} = \mathcal{C}\}^+$.*

Proof. The arguments are analogous to those in the previous proof, except that we leave out the last few lines. □

Assume that Z contains a regular special point. Choose a normal algebraic subgroup $N_2 \lhd M$ such that M is the almost direct product of N_Z and N_2. From this we obtain a decomposition $(M^{\mathrm{ad}}, \mathfrak{Y}^{\mathrm{ad}}) = (N_Z^{\mathrm{ad}}, \mathfrak{Y}_1') \times (N_2^{\mathrm{ad}}, \mathfrak{Y}_2')$. (We write $\mathfrak{Y}_1'$ and $\mathfrak{Y}_2'$ to avoid confusion with the decomposition $\mathfrak{Y}^{\mathrm{ad}} = \mathfrak{Y}_1 \times \mathfrak{Y}_2$ introduced before.)

From the remark that $N_Z^{\mathrm{ad}}(\mathbb{R})^+$ stabilizes $\mathcal{C}$ it easily follows that there exists a component Y_1' and a class $\eta' K \in G(\mathbb{A}_f)/K$ such that $S_{\eta' K}(Y_1', P) \subseteq Z$ for every point P in the image of the projection map $\mathcal{C} \to \mathfrak{Y}_2'$. Notice that this gives interesting information only if $Z(M)$ is a proper subgroup of N_Z.

§4 Totally geodesic subvarieties

4.1 Let $Z \hookrightarrow Sh_K(G, \mathfrak{X})$ be an irreducible subvariety of a Shimura variety. Choose a connected component $X \subseteq \mathfrak{X}$ and a class $\eta K \in G(\mathbb{A}_f)/K$ such that Z is contained in the image of $X \times \eta K$ in $Sh_K(G, \mathfrak{X})$. We say that Z is a totally geodesic subvariety if there is a totally geodesic subvariety $Y \subseteq X$ (in the sense of [20, Chap. 1, §14]) such that Z is the image of $Y \times \eta K$ in $Sh_K(G, \mathfrak{X})$.

We can of course express total geodesicness directly in terms of the metric on $Sh_K(G, \mathfrak{X})$, provided that we take possible singularities into account. For example, if $Sh_K(G, \mathfrak{X})$ is non-singular then $Z \hookrightarrow Sh_K(G, \mathfrak{X})$ is totally geodesic if and only if every geodesic in $Sh_K(G, \mathfrak{X})$ which is tangent to Z at a regular point $P \in Z^{\mathrm{reg}}$ is a curve in Z.

It is immediate from the definitions that subvarieties of Hodge type, and, more generally, subvarieties of the form $S_{\eta K}(Y_1, y_2)$ are totally geodesic.

4.2 Remark. The fact that we are working in an ambient space ($Sh_K(G, \mathfrak{X})$) of constant curvature implies that total geodesicness needs to be tested only at one point. To formulate this more precisely, consider an irreducible subvariety $Z \hookrightarrow Sh_K(G, \mathfrak{X})$ as above, and let $P \in Z$ be a regular point of $Sh_K(G, \mathfrak{X})$. Then Z is called totally geodesic at P if every geodesic in $Sh_K(G, \mathfrak{X})$ which is tangent to Z at P is locally a curve in Z. The remark now is that if Z is totally geodesic at one such point P then Z is totally geodesic in the sense of the previous definition. See also [5, p. 195].

Building on the results of the previous sections, we can now establish one of the main results of this paper. This result was suggested to us by D. Kazhdan.

4.3 Theorem. *Let $(G, \mathfrak{X})$ be a Shimura datum, and let K be a compact open subgroup of $G(\mathbb{A}_f)$. An algebraic subvariety $Z \hookrightarrow Sh_K(G, \mathfrak{X})$ is totally geodesic if and only if there exists a closed immersion of*

Shimura data $i\colon (M,\mathfrak{Y}) \hookrightarrow (G,\mathfrak{X})$, *a decomposition* $(M^{\mathrm{ad}},\mathfrak{Y}^{\mathrm{ad}}) = (M_1,\mathfrak{Y}_1)\times(M_2,\mathfrak{Y}_2)$, *a component* $Y_1 \subseteq \mathfrak{Y}_1$, *a point* $y_2 \in \mathfrak{Y}_2$ *and a class* $\eta K \in G(\mathbb{A}_f)/K$ *such that* $Z = S_{\eta K}(Y_1, y_2)$ *(as defined in Section 3.1).*

If Z *contains a special point, then* Z *is totally geodesic if and only if it is of Hodge type.*

Proof. It suffices to prove the theorem for K sufficiently small, so we may assume that the notations and results of Section 3 apply. Take a totally geodesic subvariety Z, and consider the inclusion $Z \subseteq S_{\eta K}(Y_1, y_2)$ as in Proposition 3.7. By (ii) of Lemma 3.3 we may assume that $Y_1 \twoheadrightarrow S_{\eta K}(Y_1,y_2)$ is a covering map. By Variant 3.9 there is a normal algebraic subgroup $H_{\zeta,\mathcal{C}} \triangleleft H_{\zeta,\mathbb{R}}$ such that

$$H_{\zeta,\mathcal{C}}(\mathbb{R})^+ = \{h \in H_\zeta(\mathbb{R}) \mid h\mathcal{C} = \mathcal{C}\}^+ . \tag{1}$$

If H' is a complement for $H_{\zeta,\mathcal{C}}$ in $H_{\zeta,\mathbb{R}}$ then we obtain a decomposition $Y_1 = W_1 \times W_2$, such that $H_{\zeta,\mathcal{C}}(\mathbb{R})^+$ acts transitively on W_1 and $H'(\mathbb{R})^+$ acts transitively on W_2.

By assumption, $\mathcal{C}$ is a complete, totally geodesic submanifold of Y_1. The group $H_{\zeta,\mathcal{C}}(\mathbb{R})^+$ therefore acts transitively on $\mathcal{C}$. It follows that

$$\mathcal{C} = W_1 \times \{w_2\} \tag{2}$$

for some $w_2 \in W_2$.

Let $H'_{w_2} \subseteq H'$ be the stabilizer subgroup of the point w_2, which is an algebraic subgroup of H'. Combining (1) and (2) we see that $H'_{w_2}(\mathbb{R})^+ = \{1\}$. On the other hand, H' is a semi-simple group over $\mathbb{R}$ and $H'_{w_2}(\mathbb{R})^+$ is a maximal compact subgroup of $H'(\mathbb{R})^+$. We conclude that $H' = \{1\}$, hence W_2 is reduced to the single point w_2. It follows that $\mathcal{C} = Y_1 \times \{y_2\}$, which proves the first statement of the theorem.

Next, suppose Z contains a special point. Since $S_{\eta K}(Y_1,y_2)$ is non-singular for K sufficiently small we may assume this special point to be regular. As remarked after Proposition 3.7, this implies that $Z = S_{\eta K}(Y_1,y_2)$ is of Hodge type. □

4.4 Corollary. *Let* $Z \hookrightarrow Sh_K(G,\mathfrak{X})$ *be a totally geodesic subvariety, then there exists a subgroup* $K' \subseteq K$ *of finite index, algebraic varieties* S_1, S_2, *a closed immersion* $g\colon S_1 \times S_2 \hookrightarrow Sh_{K'}(G,\mathfrak{X})$ *and points* $a, b \in S_2$ *such that*

1. $S_1 \times \{s_2\}$ *and* $\{s_1\} \times S_2$ *are totally geodesic subvarieties of* $Sh_{K'}(G,\mathfrak{X})$ *for every* $s_1 \in S_1$, $s_2 \in S_2$,
2. Z *is the image of* $S_1 \times \{a\}$ *under* $Sh_{(K',K)}$,
3. $S_1 \times \{b\}$, *hence also* $Sh_{(K',K)}(S_1 \times \{b\})$ *is a subvariety of Hodge type.*

Proof. The map $\varphi\colon Y_1 \times Y_2 \to Sh_K(G,\mathfrak{X})$ obtained by sending (y_1,y_2) to the class $[y_1,y_2,\eta K]$ factors through a finite morphism of algebraic varieties $\varphi'\colon \Gamma\backslash(Y_1\times Y_2) \to Sh_K(G,\mathfrak{X})$, where Γ is an arithmetic subgroup of $M^{\mathrm{ad}}(\mathbb{Q})$. There are arithmetic subgroups $\Gamma_1 \subset H_z^{\mathrm{ad}}(\mathbb{Q})$ and $\Gamma_2 \subseteq H_2^{\mathrm{ad}}(\mathbb{Q})$ such that $\Gamma_1 \times \Gamma_2$ is of finite index in Γ ([6, Cor. 8.10]). Taking $S_i = \Gamma_i\backslash Y_i$ we arrive at the corollary. □

4.5 To conclude this section, let us discuss an example. The example concerns a subvariety $S \hookrightarrow \mathsf{A}_{4d,1,n}$ ($d \geq 2$) of Shimura type, such that for a generic point $\eta \in S$, the abelian variety Y_η is simple, whereas the generic Mumford-Tate group G on S has a non-simple adjoint group. This then leads to non-trivial totally geodesic subvarieties Z which are not of Hodge type, and for which the connected algebraic monodromy group H_z is a proper subgroup of G^{der}. We also obtain a negative answer to two problems formulated in [1, Chap. X].

A sketch of a special case of the example can be found in [2], where a reference is given to Borovoi's paper [8]. However, we have not had the opportunity to read (a translation of) Borovoi's paper, so it is not clear to us to whom the example is due.

Let F be a totally real field of degree $d \geq 2$ over $\mathbb{Q}$, and write $\infty_1, \ldots, \infty_d$ for its places at infinity. Take two quaternion algebras D_1, D_2 which both have at least one invariant 0 at infinity and which moreover have "complementary" invariants at infinity, i.e., $\mathrm{inv}_{\infty_i}(D_1) = 0$ if and only if $\mathrm{inv}_{\infty_i}(D_2) = 1/2$. Then $D_1 \otimes_F D_2 \cong \mathrm{M}_2(D)$ for some other quaternion algebra D over F (using that $\mathrm{inv}_v(D_1 \otimes D_2) = \mathrm{inv}_v(D_1) + \mathrm{inv}_v(D_2)$ in $\mathbb{Q}/\mathbb{Z}$, and the fact that D_1 and D_2 have different invariants at infinity). Let $G_1 = \mathrm{Res}_{F/\mathbb{Q}} D_1^*, G_2 = \mathrm{Res}_{F/\mathbb{Q}} D_2^*$, let $V = D \oplus D$ as a $\mathbb{Q}$-vector space, and define the homomorphism

$$f \colon G_1 \times G_2 \to \mathrm{GL}(V)$$

as the composition of $G_1 \times G_2 \to \mathrm{Res}_{F/\mathbb{Q}}(\mathrm{GL}_2(D))$ and the natural map

$$\mathrm{Res}_{F/\mathbb{Q}}(\mathrm{GL}_2(D)) \hookrightarrow \mathrm{GL}(\mathrm{Res}_{F/\mathbb{Q}}(D \oplus D)) = \mathrm{GL}(V)\,.$$

Let $\mathfrak{X}_1$ be the $G_1(\mathbb{R})$-conjugacy class in $\mathrm{Hom}(\mathbb{S}, G_{1,\mathbb{R}})$ of the homomorphism h_1 given on $\mathbb{R}$-valued points by

$$\mathbb{C}^* \ni a+bi \mapsto \Big(\prod_{j \in J_1} \mathrm{Id} \times \prod_{j \in J_2} \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \Big) \in \prod_{j \in J_1} \mathbb{H}^* \times \prod_{j \in J_2} \mathrm{GL}_2(\mathbb{R}) \cong G_1(\mathbb{R})\,,$$

where $J_1 = \{j \mid \mathrm{inv}_{\infty_j}(D_1) = 1/2\}$, $J_2 = \{j \mid \mathrm{inv}_{\infty_j}(D_1) = 0\}$. Notice that $\mathfrak{X}_1$ is well-defined since all automorphisms of $\mathrm{GL}_{2,\mathbb{R}}$ are inner. Likewise we get a $G_2(\mathbb{R})$-conjugacy class $\mathfrak{X}_2$ in $\mathrm{Hom}(\mathbb{S}, G_{2,\mathbb{R}})$. One easily checks that $(G_i, \mathfrak{X}_i)$ is a Shimura datum, i.e., a pair satisfying conditions (2.1.1.1-3) of [17, Sect. 2.1]. In this way we get Shimura varieties $Sh(G_1, \mathfrak{X}_1)$ and $Sh(G_2, \mathfrak{X}_2)$. Notice that these are not of Hodge type, since their weight is not defined over $\mathbb{Q}$—see the lemma below.

Let G be the image of $G_1 \times G_2$ under f and consider the $G(\mathbb{R})$-conjugacy class $\mathfrak{X}$ in $\mathrm{Hom}(\mathbb{S}, G_\mathbb{R})$ which is the image of $\mathfrak{X}_1 \times \mathfrak{X}_2$ under the natural map $\mathrm{Hom}(\mathbb{S}, G_{1,\mathbb{R}}) \times \mathrm{Hom}(\mathbb{S}, G_{2,\mathbb{R}}) \to \mathrm{Hom}(\mathbb{S}, G_\mathbb{R})$. The pair $(G, \mathfrak{X})$ thus obtained is again a Shimura datum. Note that we have an exact sequence

$$1 \to \mathrm{Res}_{F/\mathbb{Q}}\, \mathbb{G}_{\mathrm{m},F} \xrightarrow{\Delta^-} G_1 \times G_2 \to G \to 1\,,$$

where Δ^- is (the Weil restriction of) the antidiagonal map $F^* \ni f \mapsto (f, f^{-1}) \in D_1^* \times D_2^*$. It follows from this that $\mathfrak{X}_1 \times \mathfrak{X}_2 \cong \mathfrak{X}$.

For a compact open subgroup K of $G(\mathbb{A}_f)$ let $K_i = f_i^{-1}(K)$, where f_i is the restriction of f to G_i. Then K_i is a compact open subgroup of $G_i(\mathbb{A}_f)$ and we get a morphism

$$f_{(K_1 \times K_2, K)} \colon Sh_{K_1 \times K_2}(G_1 \times G_2, \mathfrak{X}_1 \times \mathfrak{X}_2) \longrightarrow Sh_K(G, \mathfrak{X})\,.$$

We choose connected components X_1, X_2 and denote the corresponding connected Shimura varieties by $Sh(G_i, X_i)$. For K_1, K_2 sufficiently small the map $f^0_{(K_1\times K_2,K)}$ on connected Shimura varieties is finite étale.

4.6 Lemma. *The weight homomorphisms $w_i\colon \mathbb{G}_\mathrm{m} \to G_{i,\mathbb{R}}$ $(i = 1, 2)$ are not defined over $\mathbb{Q}$; the weight homomorphism $w\colon \mathbb{G}_\mathrm{m} \to G_\mathbb{R}$ is defined over $\mathbb{Q}$.*

Proof. Let $Z_i := Z(G_i)$, and $Z := Z(G)$. Then $Z_i \cong \mathrm{Res}_{F/\mathbb{Q}}\mathbb{G}_{\mathrm{m},F}$ $(i = 1, 2)$, and $Z = Z_1 \times Z_2/\mathrm{Im}(\Delta^-)$. The character group $X^*(Z_i)$ is the free abelian group on the set $\{\infty_1, \ldots, \infty_d\}$, with its natural structure of a $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-module. If $f_{i,1}, \ldots, f_{i,d}$ is the dual basis for the cocharacter group $X_*(Z_i)$, then

$$X_*(Z) = (\mathbb{Z}\cdot f_{1,1} + \cdots + \mathbb{Z}\cdot f_{1,d}) \oplus (\mathbb{Z}\cdot f_{2,1} + \cdots + \mathbb{Z}\cdot f_{2,d})/\langle f_{1,i} - f_{2,i}; 1 \le i \le d\rangle\,.$$

As is always the case for Shimura data, the weight homomorphisms w_i (resp. w) take values in the center $Z_{i,\mathbb{R}}$ (resp. $Z_\mathbb{R}$). It is immediate from the definitions that

$$w_1 = \sum_{j\in J_2} f_{1,j} \in X_*(Z_1)\,, \quad w_2 = \sum_{j\in J_1} f_{2,j} \in X_*(Z_2)\,,$$

which are not $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-invariant since J_1 and J_2 are both nonempty. The weight w is given by

$$w = w_1 \oplus w_2 \bmod \langle f_{1,i} - f_{2,i}\rangle = f_{1,1} + \cdots + f_{1,d} \bmod \langle f_{1,i} - f_{2,i}\rangle = f_{2,1} + \cdots + f_{2,d} \bmod \langle f_{1,i} - f_{2,i}\rangle\,,$$

and this clearly is Galois invariant. □

By considering the representation of $G_\mathbb{R}$ on $V_\mathbb{R}$ one sees that there does not exist a symplectic form Ψ on V such that G acts through symplectic similitudes. Essentially the problem is that the center of G is "too large". Therefore, we introduce the algebraic subgroup $G' = w(\mathbb{G}_\mathrm{m})\cdot G^\mathrm{der} \subseteq G$, which, by the lemma, is defined over $\mathbb{Q}$. All $h_x\colon \mathbb{S} \to G_\mathbb{R}$ for $x \in \mathfrak{X}$ factor through $G'_\mathbb{R}$, and we have a closed immersion of Shimura data $(G', \mathfrak{X}) \hookrightarrow (G, \mathfrak{X})$.

The lemma shows that $(G', \mathfrak{X})$ satisfies condition (2.1.1.4) of [17, Sect. 2.1]. It also satisfies loc. cit., condition (2.1.1.5), as one easily verifies. Furthermore, for $x \in \mathfrak{X}$ the representation h_x on V is of type $(-1, 0)+(0, -1)$. From [17, Prop. 2.3.2] it now follows that there exists a symplectic form Ψ on V such that the inclusion $G' \hookrightarrow \mathrm{GL}(V)$ induces a morphism of Shimura data $i\colon (G', \mathfrak{X}) \hookrightarrow (\mathrm{CSp}(V, \Psi), \mathfrak{H}^\pm_{4d})$. Here we identify the Siegel double space $\mathfrak{H}^\pm_{4d}$ as the space of $\mathbb{R}$-Hodge structures $h\colon \mathbb{S} \to \mathrm{GL}(V)$ of type $(-1, 0) + (0, -1)$ such that $\pm\Psi$ is a polarization. This shows that $Sh(G', \mathfrak{X})$ is a Shimura variety of Hodge type.

For a compact open subgroup $K \subseteq G(\mathbb{A}_f)$, write $K' = G'(\mathbb{A}_f) \cap K$. For K sufficiently small we get a "universal" family $\alpha\colon (Y, \lambda, \theta) \to Sh^0_{K'}(G', \mathfrak{X})$ of $4d$-dimensional principally polarized abelian varieties with a level K'-structure. The morphism $f^0_{(K_1\times K_2,K)}$ on connected Shimura varieties factors through $Sh^0_{K'}(G', \mathfrak{X})$. We can choose a point $x_2 \in X_2$ such that the subvariety

$$Z_{x_2} = Sh(f)(Sh_{K_1}(G_1, X_1) \times [x_2, eK_2]) = S_{eK'}(X_1, x_2) \subset Sh_{K'}(G', X)$$

is not contained in the locus Σ (applying the discussion of Section 1.2 to $S = Sh_{K'}(G', X)$ and the natural VHS with local system $\mathrm{R}^1\alpha_*\mathbb{Z}_Y$). The generic fibre Y_η on $Z = Z_{x_2}$ has Mumford-Tate group G', and the representation of $\mathrm{MT}(Y_\eta)$ on $\mathrm{H}^1(Y_\eta(\mathbb{C}), \mathbb{Q})$ is isomorphic to $G' \hookrightarrow \mathrm{GL}(V)$. In particular, V being an irreducible G'-module, Y_η is simple. On the other hand, it is clear that the connected algebraic monodromy group of the restricted family (Y, λ, θ) over Z is contained in $f(G_2)^{\mathrm{der}}$, so it is strictly contained in $(G')^{\mathrm{der}} = G^{\mathrm{der}}$.

As remarked by André ([2], footnote on p. 13) the example contradicts the conjectural statement IX, 3.1.6 in [1]. We claim that it also gives a negative answer to op. cit., Chap. X, Problems 2 and 3.

Loc. cit., Problem 2 is essentially the following. Consider a subvariety $Z \hookrightarrow \mathsf{A}_{g,1,n} \otimes \mathbb{C}$ satisfying (i) $\dim(Z) = 1$, (ii) the generic fibre in the family of abelian varieties over Z is simple, (iii) there are infinitely many points on Z which lie on a proper subvariety of Hodge type. Does it follow that Z is of Hodge type? We see that the answer is negative in general: in the above example we choose D_1 and D_2 such that $\#J_2 = 1$, which implies that $\dim(Z_{x_2}) = 1$. As we have seen, Z_{x_2} satisfies condition (ii) and it is not of Hodge type. Finally, for all special points $x_1 \in X_1$, the point $Sh(f)([x_1, eK_1] \times [x_2, eK_2]) \in Z_{x_2}$ lies on a proper subvariety of Hodge type.

A special case of loc. cit., Problem 3, is the following question. Consider a subvariety $Z \hookrightarrow \mathsf{A}_{g,1,n} \otimes \mathbb{C}$ satisfying conditions (i) and (ii) and also satisfying (iv) there are infinitely many points on Z such that the corresponding abelian varieties are all isogenous. Does it follow that Z is of Hodge type? Again, the answer is negative. The example is the same as above; for (iv) we only have to remark that for a fixed $x_1 \in X_1$, the fibres over the points $Sh(f)([g_1 \cdot x_1, eK_1] \times [x_2, eK_2]) \in Z_{x_2}$ with $g_1 \in G_1(\mathbb{Q})$ are all isogenous.

§5 Serre-Tate group structures over $\mathbb{C}$

5.1 In this section we reformulate Theorem 4.3 in terms of what we call a "Serre-Tate group structure" over $\mathbb{C}$. This is to be compared to classical Serre-Tate theory (see [21]) which is at the basis of the "linearity property" studied in the second part of the paper. The formulation of "total geodesicness" in terms of a formal group structure reveals a close analogy between the theory over $\mathbb{C}$ studied so far and the theory in mixed characteristics which is the main topic of Part II.

5.2 Consider a Shimura datum $(G, \mathfrak{X})$ and let $K \subset G(\mathbb{A}_f)$ be a compact open subgroup. Given a connected component $X \subseteq \mathfrak{X}$ and a class $\eta K \in G(\mathbb{A}_f)/K$, we have a uniformization map

$$u\colon\quad \begin{aligned} X &\longrightarrow Sh_K(G, \mathfrak{X}) \\ x &\longmapsto [x, \eta K] \quad . \end{aligned}$$

We assume that K is small enough such that the map u from X to its image $Sh^0 \subseteq Sh_K(G, \mathfrak{X})$ is a topological covering.

Write $X \hookrightarrow \check{X}$ for the Borel embedding of the hermitian symmetric domain X into its compact dual. Choose a point $x \in X$, let $\theta = \theta_x := \mathrm{Inn}(h_x(i))$ denote the associated Cartan involution, and write $\mathcal{G}^{(\theta)}$ for the corresponding compact real form of $G^{\mathrm{ad}}_{\mathbb{R}}$. The domain $\check{X}$ is a homogenous space under $\mathcal{G}^{(\theta)}(\mathbb{R})$. One also has a realization

$$\check{X} = G^{\mathrm{ad}}(\mathbb{C})/P_x(\mathbb{C}) ,$$

where $P_x \subset G^{\mathrm{ad}}_{\mathbb{C}}$ is the stabilizer of the point x, which is a parabolic subgroup. It should be noted that if X has positive dimension then $G^{\mathrm{ad}}(\mathbb{C})$ does not act on $\check{X}$ by isometries.

Let $K_\infty = P_x(\mathbb{C}) \cap G^{\mathrm{ad}}(\mathbb{R})$ denote the stabilizer of the point $x \in X$ inside $G^{\mathrm{ad}}_{\mathbb{R}}$, and write $\mathfrak{g} = \mathrm{Lie}(G^{\mathrm{ad}}_{\mathbb{R}})$ and $\mathfrak{k} = \mathrm{Lie}(K_\infty)$. The Hodge decomposition of $\mathfrak{g}_{\mathbb{C}}$ with respect to $\mathrm{Ad} \circ h_x$ is of the form

$$\mathfrak{g}_{\mathbb{C}} = \mathfrak{u}^- + \mathfrak{k}_{\mathbb{C}} + \mathfrak{u}^+ ,$$

where $\mathfrak{k}_{\mathbb{C}} + \mathfrak{u}^+ = \mathrm{Lie}(P_x)$. If $P_x^- \subset G^{\mathrm{ad}}_{\mathbb{C}}$ is the parabolic subgroup with Lie algebra $\mathfrak{k}_{\mathbb{C}} + \mathfrak{u}^-$ then P_x and P_x^- are opposite parabolic subgroups with common Levi factor $K_\infty \otimes \mathbb{C}$.

Write U_x^- for the unipotent radical of P_x^- (with $\mathrm{Lie}(U_x^-) = \mathfrak{u}^-$). The natural map $U_x^-(\mathbb{C}) \to \check{X}$ gives an isomorphism of $U_x^-(\mathbb{C})$ onto its image $\mathcal{U} \subset \check{X}$, which is the complement of an ample divisor $D \subset \check{X}$.

Via the uniformization map $u\colon X \twoheadrightarrow Sh^0 \subseteq Sh_K(G,\mathfrak{X})$ we can now describe Sh^0 "locally at the point $y = u(x)$ as the germ of U_x^- at the identity element. To make this more precise, we consider formal completions: write $\mathfrak{U}_x$ for the completion of U_x^- at the identity element, and let $\mathfrak{Sh}_y$ denote the formal completion of Sh^0 at the point $y = u(x)$. Then $\mathfrak{U}_x$ is isomorphic to the formal completion of $\check{X}$ at x, and u induces an isomorphism $\hat{u}_x\colon \mathfrak{U}_x \xrightarrow{\sim} \mathfrak{Sh}_y$. In this way the formal completion $\mathfrak{Sh}_y$ inherits the structure of a formal group isomorphic to $\widehat{\mathbb{G}}_{\mathrm{a}}^d$ ($d = \dim Sh_K(G,\mathfrak{X}) = \dim X$). One readily verifies that this structure does not depend on the choice of a point $x \in X$ with $u(x) = y$.

Our results in this paper and in Part II will show that this structure of a formal group on $\mathfrak{Sh}_y$ is analogous to the Serre-Tate group structure on the formal deformation space of an ordinary abelian variety in characteristic $p > 0$. Therefore we will refer to the group structure on $\mathfrak{Sh}_y$ just defined as the "Serre-Tate group structure". Notice, however, that in this case we are working with formal vector groups (i.e., formal groups isomorphic to a power of $\widehat{\mathbb{G}}_{\mathrm{a}}$), whereas the theory in mixed characteristics produces formal tori.

5.3 Definition. Let $Sh^0 \subseteq Sh_K(G,\mathfrak{X})$ be an irreducible component as before, and let $Z \hookrightarrow Sh^0$ be an irreducible algebraic subvariety. We say that Z is formally linear at the point y if the formal completion $\mathfrak{Z}_y = Z_{/\{y\}} \hookrightarrow \mathfrak{Sh}_y$ is a formal vector subgroup of $\mathfrak{Sh}_y$.

5.4 Proposition. *Let $Sh^0 \subseteq Sh_K(G,\mathfrak{X})$ be an irreducible component, and suppose that the uniformization map $u\colon X \twoheadrightarrow Sh^0$ is a topological covering. Let $Z \hookrightarrow Sh^0$ be an irreducible algebraic subvariety, and let $y \in Z$. Then Z is totally geodesic at y if and only if Z is formally linear at y.*

Proof. In the "only if" direction this is not difficult: if Z is totally geodesic at y then, as pointed out in 4.2, Z is totally geodesic everywhere, hence Z is of the form $Z = S_{\eta K}(Y_1, y_2)$ as in Theorem 4.3. From this it follows directly that Z is formally linear at y.

For the converse, suppose that Z is formally linear at the point y. With some minor exceptions, we return to the notations of Section 3. In particular, we write $S = S_{\eta K}(Y_M) \hookrightarrow Sh^0$ for the smallest subvariety of Hodge type containing Z. Since in proving the proposition we may pass to a higher level, we may assume that the conditions in 3.2 are satisfied, and that the algebraic monodromy group H_ζ of the local system $n^*\mathcal{V}(\xi)$ over the normalization Z^n is connected. This last assumption is only made to avoid some clumsy formulations.

Consider the inclusion $Z \hookrightarrow S_{\eta K}(Y_1, y_2)$ as in Proposition 3.7 (resulting from the decomposition $M^{\mathrm{ad}} = H_\zeta^{\mathrm{ad}} \times H_2^{\mathrm{ad}}$). Choose $\mathcal{C} \subseteq \widetilde{Z}$ as in 3.4, and let $x \in \mathcal{C}$ be a point mapping to y. We are back in the situation of the diagram in 3.4, with the additional information that $\mathcal{C} \subseteq Y_1 \times \{y_2\}$.

The group U_x^- is connected unipotent and abelian (since $[\mathfrak{u}^-, \mathfrak{u}^-] \subset \mathfrak{g}_{\mathbb{C}}$ is of type $(-2,2)$, hence zero), so we have $U_x^- \cong \mathbb{G}_{\mathrm{a}}^d$, for $d = \dim(X)$. Let $t_1, \ldots, t_d$ be the pull-backs to U_x^- of the standard coordinates on $\mathbb{G}_{\mathrm{a}}^d$. The t_i, viewed as functions on $\mathcal{U} \subset \check{X}$, extend to global sections $t_i \in \Gamma(\check{X}, \mathcal{L})$, where $\mathcal{L} = \mathcal{O}_{\check{X}}(k \cdot D)$ for a suitable $k \geq 0$. We can choose this k such that there is an action of $G_{\mathbb{C}}^{\mathrm{ad}}$ on the line bundle $\mathcal{L}$, making it a $G_{\mathbb{C}}^{\mathrm{ad}}$-bundle over $\check{X}$. (This is probably well-known to experts. The point is that $\mathcal{U}$ is the big open cell $\mathcal{U} = R_{\mathrm{u}}(B^-) \cdot P_x$ in the Bruhat decomposition relative to a Borel subgroup $B \subset P_x$. If Δ is the corresponding basis of the root system, then P_x is a standard parabolic subgroup corresponding to a subset $I \subset \Delta$. The divisor D has a number of components $D_\alpha = \{R_{\mathrm{u}}(B^-) \cdot s_\alpha \cdot P_x\}^-$, one for each $\alpha \in \Delta \setminus I$. Write ω_α for the fundamental weight corresponding to α, and choose k such that $\lambda := k \cdot \sum_{\alpha \in \Delta \setminus I} \omega_\alpha \in X^*(T)$, where T is the maximal torus $B \cap B^-$. Then λ gives a character of P_x and it follows from the results in [4] that $\mathcal{O}_{\check{X}}(k \cdot D)$ is the associated line bundle $\mathcal{L}(\lambda)$ on $\check{X} = G_{\mathbb{C}}^{\mathrm{ad}}/P_x$.)

Define

$$I = \{s \in \Gamma(\check{X}, \mathcal{L}) \mid s_{|\mathcal{C}} = 0\},$$

and write $V(I)$ for the zero locus of I. We claim that $\mathcal{C}$ is an irreducible component of $V(I) \cap X$. To see this we argue as follows. Let $\tau_1, \ldots, \tau_d$ be the coordinates on $\mathfrak{U}_x$ induced by the t_i. The assumption that Z is formally linear at the point y implies that the formal completion

$$\mathcal{C}_{/\{x\}} \longrightarrow \check{X}_{/\{x\}} \cong \mathfrak{U}_x$$

is defined by a number of linear equations $\sum m_i \cdot \tau_i = 0$. The corresponding sections $\sum m_i \cdot t_i \in \Gamma(\check{X}, \mathcal{L})$ are elements of I, since they define holomorphic functions on $\mathcal{C}$ with vanishing Taylor expansion at x. The claim readily follows from this.

Recall that we have chosen a representation ξ of G as in 3.2. We have a monodromy action $\rho_S \colon \pi_1(S, x) \to \xi(M)(\mathbb{Q})$, and since $\mathcal{L}$ is a G^{ad}-bundle over $\check{X}$ this induces an action of $\pi_1(Z^n, \zeta)$ on $\Gamma(\check{X}, \mathcal{L})$. It follows from Lemma 3.5 that the subspace $I \subseteq \Gamma(\check{X}, \mathcal{L})$ is stable under $\pi_1(Z^n, \zeta)$. Since the image of $\pi_1(Z^n, \zeta)$ is Zariski dense in $H_{\zeta,\mathbb{R}}$ (by definition of H_ζ it is dense in $H_{\zeta,\mathbb{Q}}$, and this implies that it is dense in $H_{\zeta,\mathbb{R}}$), this implies that $V(I) \cap X$ is stable under the action of $H_\zeta(\mathbb{R})$. It follows that $(y_1, b) \in \mathcal{C} \Rightarrow Y_1 \times \{b\} \subseteq \mathcal{C}$ for $y_1 \in Y_1$, $b \in Y_2$, and since we already know that $\mathcal{C} \subseteq Y_1 \times \{y_2\}$ for some point $y_2 \in Y_2$, this proves that Z is a subvariety of the form $S_{\eta K}(Y_1, y_2)$. □

5.5 Corollary. *An irreducible algebraic subvariety $Z \hookrightarrow Sh_K(G, \mathfrak{X})$ is of Hodge type if and only if Z is formally linear at some point and Z contains a special point.*

§6 Non-rigid families of abelian varieties

6.1 In his paper [3], Arakelov proved that, given a complete and non-singular curve B over $\mathbb{C}$, a finite set of points $S \subseteq B$ and an integer $g \geq 2$, the set

$$\left\{ \begin{array}{l} \text{isomorphism classes of non-constant families of} \\ \text{non-singular irreducible curves of genus } g \text{ over } B \setminus S \end{array} \right\}$$

is finite. (For $S = \emptyset$ this was done by Paršin.) One of the main steps in the proof was to show that if X is such a non-constant family of non-singular curves of genus g over $B \setminus S$, then X does not have non-trivial deformations.

In [19], Faltings gave an example showing that in general the analogous statement for abelian varieties is not true. His example concerns a non-rigid family of abelian eightfolds. After that, several people came up with related results. As for the non-rigid families of abelian varieties, Masa-Hiko Saito obtained in [26] a classification of the endomorphism algebras of the underlying local systems. In particular, he determined for which (relative) dimensions there exist such non-rigid families (without isotrivial factors).

Using the notations and results discussed before, we can add to Saito's work. We describe the non-rigid families of abelian varieties in terms of the corresponding subvarieties of the moduli space and we "explain" the non-rigidity geometrically. Before we do so, let us first discuss the problem in a rather general setting.

6.2 Suppose we are given a polarized $\mathbb{Z}$-VHS $\mathcal{V} = (\mathcal{V}_{\mathbb{Z}}, \mathcal{F}^{\bullet}, \mathcal{Q})$ of pure weight n over a non-singular, irreducible, complex algebraic variety Z. We can ask if there are non-trivial deformations of $\mathcal{V}$, fixing the base space Z. We are in fact most interested in the infinitesimal deformations. To give this a precise meaning one has to set up some theory.

Part of the structure of $\mathcal{V}$ is discrete, and therefore cannot vary continuously. Specifically, let $\mathcal{T}$ be the set of equivalence classes of 4-tuples $(V_{\mathbb{Z}}, Q, \pi, \rho)$, where $V_{\mathbb{Z}}$ is a free $\mathbb{Z}$-module of finite rank, $Q\colon V_{\mathbb{Z}} \times V_{\mathbb{Z}} \to \mathbb{Z}$ is a bilinear form, π is a group and $\rho\colon \pi \to \mathrm{Aut}(V_{\mathbb{Z}}, Q)$ is a homomorphism. Two such 4-tuples $(V_{\mathbb{Z}}, Q, \pi, \rho)$ and $(V'_{\mathbb{Z}}, Q', \pi', \rho')$ are said to be equivalent if there exist isomorphisms $\alpha\colon (V_{\mathbb{Z}}, Q) \xrightarrow{\sim} (V'_{\mathbb{Z}}, Q')$ and $\beta\colon \pi \xrightarrow{\sim} \pi'$ such that $\alpha_* \circ \rho = \rho' \circ \beta$. For $z \in Z$ we get such a 4-tuple by taking $(V_{\mathbb{Z}}, Q) = (\mathcal{V}_z, \mathcal{Q}_z)$, $\pi = \pi_1(Z, z)$ and $\rho\colon \pi_1(Z, z) \to \mathrm{Aut}(V, Q)$ the monodromy representation. The class of this 4-tuple in $\mathcal{T}$ does not depend on the choice of z, and therefore we get a well-determined element $\tau(\mathcal{V}) \in \mathcal{T}$ associated to $\mathcal{V}$.

In Peters' paper [25] it is shown that the set

$$\mathbb{Z}\text{-}\mathsf{VHS}_{\tau}(Z) = \left\{ \begin{array}{l} \text{isomorphism classes of polarized } \mathbb{Z}\text{-VHS} \\ \mathcal{V} \text{ over } Z \text{ with } \tau(\mathcal{V}) = \tau \in \mathcal{T} \end{array} \right\}$$

has a natural structure of an analytic variety such that the tangent space to the class $[\mathcal{V}]$ is isomorphic to $(E^Q \otimes_{\mathbb{Q}} \mathbb{C})^{-1,1}$. Here $E = \mathrm{H}^0(Z, \mathcal{E}nd(\mathcal{V}_{\mathbb{Q}}))$, the algebra of global (flat) endomorphisms of the local system $\mathcal{V}_{\mathbb{Q}}$, and

$$E^Q = \{e \in E \mid Q(ev, w) + Q(v, ew) = 0 \quad \text{for all sections } v, w \text{ of } \mathcal{V}_{\mathbb{Q}}\}$$

is the subspace of elements of E which are skew-symmetric with respect to Q. We conclude that the polarized $\mathbb{Z}$-VHS $\mathcal{V}$ over Z is rigid (i.e., it has no infinitesimal deformations over Z) if and only if $(E^Q \otimes_{\mathbb{Q}} \mathbb{C})^{-1,1} = 0$. The problem that we are interested in is to describe, or to classify, the polarizable $\mathbb{Z}$-VHS $\mathcal{V}$ over Z such that $\mathcal{V}$ is non-rigid over some finite covering of Z.

6.3 Let $z \in Z$ be a Hodge-generic point, and write $M = \mathrm{MT}_z$, and $(V, Q) = (\mathcal{V}_{\mathbb{Q},z}, Q_z)$. The Hodge group $\mathrm{Hg} = \mathrm{Hg}_z$ is defined as the intersection of MT_z and $\mathrm{SL}(V)$. Since we allow finite coverings of Z we may assume the algebraic monodromy group to be connected. Then we have algebraic groups

$$H_z \lhd M^{\mathrm{der}} = \mathrm{Hg}^{\mathrm{der}} \subseteq \mathrm{Aut}(V, Q)\,,$$

and E^Q is just the space of H_z-invariants in $\mathrm{End}^Q(V) := \{e \in \mathrm{End}(V) \mid \forall v, w \in V : Q(ev, w) + Q(v, ew) = 0\}$.

From now on we assume the VHS $\mathcal{V}$ to be of type $(-1,0)+(0,-1)$, which means that it corresponds to a family of abelian varieties over Z. We will make free use of the correspondence between polarizable $\mathbb{Z}$-VHS of type $(-1,0)+(0,-1)$ and families of abelian varieties; in particular, we say that a family of abelian varieties is non-rigid if the corresponding VHS is non-rigid.

One of the main advantages of restricting our attention to abelian varieties is that in this case the endomorphism algebra $E = \mathrm{End}(V)^{H_z}$ is of type $(-1,1)+(0,0)+(1,-1)$. Therefore, the family is rigid if and only if E^Q is purely of type $(0,0)$, which is equivalent to saying that H_z and Hg have the same invariants in $\mathrm{End}^Q(V)$. We will prove the following statement.

6.4 Theorem. *Let $f\colon X \to Z$ be a principally polarized abelian scheme over a normal irreducible complex algebraic variety Z, and assume that X admits a Jacobi level n structure θ for some $n \geq 3$. Let $\varphi_f\colon Z \to \mathsf{A}_{g,1,n} \otimes \mathbb{C}$ be the corresponding morphism of Z into the moduli space. Then there exists a closed immersion of Shimura data $i\colon (N, \mathfrak{Y}_N) \hookrightarrow (\mathrm{CSp}_{2g}, \mathfrak{H}_g^{\pm})$, a decomposition $(N^{\mathrm{ad}}, \mathfrak{Y}_N^{\mathrm{ad}}) = (N_1, \mathfrak{Y}_1) \times (N_2, \mathfrak{Y}_2)$ and a diagram*

$$\begin{array}{ccccccc} & & S_1 \times \{a\} & \longrightarrow & S_1 \times S_2 & \longrightarrow & \mathsf{A}_{g,1,K'} \otimes \mathbb{C} \\ & & {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle g'} & & \downarrow \\ Z & \xrightarrow{\varphi_f} & S = S_{\eta K}(Y_1, v) & \hookrightarrow & S_{\eta K}(Y_1 \times Y_2) & \hookrightarrow & \mathsf{A}_{g,1,n} \otimes \mathbb{C} \end{array}$$

such that there are natural isomorphisms

$$T_v Y_2 \cong T_a S_2 \xrightarrow{\sim} T_{[g^*\mathcal{W}]}(\mathbb{Z}\text{-}\mathsf{VHS}(S_1)) \xleftarrow{\sim} T_{[\mathcal{W}]}(\mathbb{Z}\text{-}\mathsf{VHS}(S)) \xrightarrow{\sim} T_{[\mathcal{V}]}(\mathbb{Z}\text{-}\mathsf{VHS}(Z))\,.$$

Here we write $\mathcal{V}$, $\mathcal{W}$ and $\mathcal{U}$ for the VHS corresponding to the first homology of the abelian schemes over Z, S and $S_1 \times S_2$ respectively, and the map $T_a S_2 \xrightarrow{\sim} T_{[g^*\mathcal{W}]}(\mathbb{Z}\text{-}\mathsf{VHS}_{\tau(g^*\mathcal{W})}(S_1))$ is the map on tangent spaces induced by the map $S_2 \to \mathbb{Z}\text{-}\mathsf{VHS}_{\tau(g^*\mathcal{W})}(S_1)$ sending $s_2 \in S_2$ to the class of $\mathcal{U}_{|S_1 \times \{s_2\}}$ in $\mathbb{Z}\text{-}\mathsf{VHS}_{\tau(g^*\mathcal{W})}(S_1)$.

Proof. Write $\lambda\colon X \to X^t$ (X^t denoting the dual abelian scheme) for the given polarization and $\mathcal{V}$ for the polarized $\mathbb{Z}$-VHS over Z corresponding to X. Without loss of generality, we may assume that the algebraic monodromy group is connected. We keep the notations $z \in Z$, H_z, M, Hg and (V, Q) as above.

Fix an integer $n \geq 3$. Possibly after passing to a finite covering of Z we can choose a Jacobi level n structure on X over Z. The family (X, λ) plus the choice of this level structure corresponds to a morphism $\varphi_f \colon Z \to \mathsf{A}_{g,1,n}(\mathbb{C})$. The fact that the generic Mumford-Tate group is M means that φ_f maps Z into a subvariety of Hodge type $S_{\eta K}(Y_M)$ (with $K \subset \mathrm{CSp}_{2g}(\mathbb{A}_f)$ the compact open subgroup corresponding to level n structures).

Recall the decomposition of Shimura data $(M^{\mathrm{ad}}, \mathfrak{Y}_M^{\mathrm{ad}}) = (H_\zeta^{\mathrm{ad}}, \mathfrak{Y}_1) \times (H_2^{\mathrm{ad}}, \mathfrak{Y}_2)$ that was introduced in Section 3.6, where $H_\zeta \lhd M$ is the connected algebraic monodromy group of the family $X \to Z$. It follows from Proposition 3.7 that $\varphi_f(Z) \subseteq S_{\eta K}(Y_1, y_2)$ for some component $Y_1 \subseteq \mathfrak{Y}_1$ and a point $y_2 \in \mathfrak{Y}_2$. Let $\mathcal{W}$ be the $\mathbb{Z}$-VHS corresponding to the universal family over $S = S_{\eta K}(Y_1, y_2)$, then we can identify the fibres $\mathcal{V}_z$ and $\mathcal{W}_{\varphi_f(z)}$, and the monodromy representation of $\mathcal{V}$ factors through that of $\mathcal{W}$. Pulling back by φ_f defines a map $\varphi_f^* \colon \mathbb{Z}\text{-}\mathsf{VHS}_{\tau(\mathcal{W})}(S) \to \mathbb{Z}\text{-}\mathsf{VHS}_{\tau(\mathcal{V})}(Z)$ which induces an isomorphism on the tangent spaces at $[\mathcal{W}]$ and $[\mathcal{V}]$ respectively. Therefore, it remains to explain the non-rigidity of the family over S. In order to keep the notations as clear as possible, we assume from now on that $Z = S = S_{\eta K}(Y_1, y_2)$ and $\mathcal{V} = \mathcal{W}$.

The basic idea now becomes apparent: Z is totally geodesic, and after Corollary 4.4 there is a product variety $S_1 \times S_2$ covering $S_{\eta K}(Y_1 \times Y_2)$, such that Z is the image of $S_1 \times \{a\}$ for some $a \in S_2$. If Y_2 (hence also S_2) is not reduced to a single point then we can vary the point $a \in S_2$, and this gives global deformations of the VHS over $S_1 \times \{a\}$. However, if Y_2 is a single point (which happens if Z is of Hodge type) then this idea does not seem to work. It does, but in general we first have to replace $(M, \mathfrak{Y}_M)$ by a “larger” Shimura datum $(N, \mathfrak{Y}_N)$. We do this as follows.

Consider the algebraic group $C = C_{\mathrm{Sp}(V,Q)}(H_z)$, the centralizer of H_z inside $\mathrm{Sp}(V, Q)$. Its connected component C^0 is a reductive subgroup of $\mathrm{Sp}(V, Q)$. Notice that $E^Q = \mathrm{Lie}(C)$. Similarly, $(E^Q)^{0,0}$ is isomorphic to the Lie algebra of $C_{\mathrm{Sp}(V,Q)}(\mathrm{Hg})$, so the fact that E^Q is not purely of type $(0,0)$ is equivalent to saying that $C_{\mathrm{Sp}(V,Q)}(\mathrm{Hg})^0 \subsetneq C^0 = C_{\mathrm{Sp}(V,Q)}(H_z)^0$.

The reductive group C^0 is the almost direct product of its center Z_{C^0} and a number of $\mathbb{Q}$-simple semi-simple factors C_i^0. Let C_c be the product of the factors C_i^0 for which $C_i^0(\mathbb{R})$ is compact, and let C' be the product of Z_C^0 and the factors C_i^0 which are not compact over $\mathbb{R}$. The intersection $H_z \cap C^0$ is contained in the center of H_z and is therefore finite, so $H_z \cdot C^0 \subseteq \mathrm{Sp}(V, Q)$ is the almost direct product of H_z and C^0. Clearly, $M \subseteq \mathbb{G}_{\mathrm{m}} \cdot H_z \cdot C^0 \subseteq \mathrm{CSp}(V, Q)$ (using that H_z is normal in M), and it follows from [17, 1.1.15] that $\mathrm{Inn}(h(i))$ is a Cartan involution of $H_z \cdot C^0$ (where $h \colon \mathbb{S} \to M_{\mathbb{R}} \subseteq \mathrm{CSp}(V, Q)_{\mathbb{R}}$ is the homomorphism giving the Hodge structure on $V = \mathcal{V}_{\mathbb{Q},z}$). The composite map

$$\mathbb{S} \to (\mathbb{G}_{\mathrm{m}} \cdot H_z \cdot C^0)_{\mathbb{R}} \xrightarrow{\mathrm{ad}} H_z^{\mathrm{ad}} \times C_c^{\mathrm{ad}} \times (C')^{\mathrm{ad}} \xrightarrow{\mathrm{pr}} C_c^{\mathrm{ad}}$$

must therefore be trivial, hence h factors through $\mathbb{G}_{\mathrm{m}} \cdot H_z \cdot C'$.

Let $N = \mathbb{G}_{\mathrm{m}} \cdot H_z \cdot C'$, and let $\mathfrak{Y}_N$ be the $N(\mathbb{R})$-conjugacy class of h. The above arguments show that $(N, \mathfrak{Y}_N)$ is a Shimura datum. We have a decomposition $(N^{\mathrm{ad}}, \mathfrak{Y}_N^{\mathrm{ad}}) = (H_z^{\mathrm{ad}}, \mathfrak{Y}_1) \times (C'^{\mathrm{ad}}, \mathfrak{Y}_{C'})$, and $S = S_{\eta K}(Y_1, y_2)$, which we can also write as $S_{\eta K}(Y_1, v)$ for a point $v \in \mathfrak{Y}_{C'}$. Notice that if there are non-trivial infinitesimal deformations, then $\mathfrak{Y}_{C'}$ is not reduced to a single point, or, equivalently: C' has non-trivial semi-simple factors. This follows from the fact that $\mathrm{Lie}(C)$ is not purely of type $(0,0)$. Alternatively: C' being a torus would contradict the above remark that $C_{\mathrm{Sp}(V,Q)}(\mathrm{Hg})^0 \subsetneq C^0$, since $\mathrm{Hg} \subseteq H_z \cdot C'$.

This brings us to a situation where we can apply Corollary 4.4. We have a commutative diagram

$$\begin{array}{ccc} S_1 \times \{a\} & \hookrightarrow & S_1 \times S_2 \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle g'} \\ S = S_{\eta K}(Y_1, v) & \hookrightarrow & S_{\eta K}(Y_1 \times Y_{C'}) \end{array}$$

where g and g' are finite surjective morphisms, and where $S_1 = Y_1/\Gamma_1$, $S_2 = Y_{C'}/\Gamma_2$ for some arithmetic subgroups $\Gamma_1 \subset H_z^{\mathrm{ad}}(\mathbb{Q})$ and $\Gamma_2 \subset C'^{\,\mathrm{ad}}(\mathbb{Q})$. We may take Γ_1 and Γ_2 small enough such that S_1 and S_2 are non-singular.

The morphism $g^*\colon \mathbb{Z}\text{-}\mathsf{VHS}_{\tau(\mathcal{W})}(S) \to \mathbb{Z}\text{-}\mathsf{VHS}_{\tau(g^*\mathcal{W})}(S_1)$ induces an isomorphism on tangent spaces at $\mathcal{W}$ and $g^*\mathcal{W}$ respectively; this follows from the description of the tangent space given above and the fact that both the generic Mumford-Tate group and the connected algebraic monodromy group on S and S_1 are the same.

Varying the point $a \in S_2$ then gives global deformations of the VHS $g^*\mathcal{W}$ over S_1. We remark that this indeed "explains all deformations": we have seen that the tangent space to $\mathbb{Z}\text{-}\mathsf{VHS}_\tau(S)$ at the point $[\mathcal{V}]$ is isomorphic to $\mathrm{Lie}(C)^{-1,1}$, and since $\mathrm{Lie}(C_c)$ is purely of type $(0,0)$ this is equal to $\mathrm{Lie}(C')^{-1,1}$. Our remark then follows from the fact that there are natural isomorphism $T_a S_2 \cong T_v Y_{C'} \xrightarrow{\sim} \mathrm{Lie}(C')^{-1,1}$ (where $Y_{C'} \ni v \mapsto a \in S_2$), which is an infinitesimal version of the correspondence

$$\text{varying the point } v \in Y_{C'} \quad \rightsquigarrow \quad \text{deformations of the VHS } g^*\mathcal{V} \text{ over } S_1\,,$$

see [25, Sect. 1 and 2] (notice that in our case the horizontal tangent bundle to the period domain $Y_{C'}$ is equal to the full tangent bundle). □

6.5 Remarks. (i) Inspection of the proof shows that the same result is valid for arbitrary VHS, provided that the Lie algebra of the centralizer C is of type $(-1,1)+(0,0)+(1,-1)$, for this is what we need to conclude that $(N, \mathfrak{Y}_N)$ is a Shimura datum.

(ii) Clearly, non-rigidity is interesting only if the family of abelian varieties over Z is not isotrivial. To exclude this possibility, we can first reduce (cf. [26, §3]) to the case where V is isotypic and non-trivial as H_z-module.

(iii) The assumption that we have a *principal* polarization is not essential; it was included only to make the result easier to state.

6.6 Corollary. *Let X be a simple abelian variety over $\mathbb{C}$ with $\dim(X) \leq 7$. Then $\mathrm{MT}(X)^{\mathrm{ad}}$ is either trivial or it is a $\mathbb{Q}$-simple algebraic group.*

Proof. This now follows immediately from the fact that there are no (non-trivial) non-rigid families of abelian varieties of relative dimension ≤ 7, see [26, Cor. 8.4]. □

References

[1] Y. ANDRÉ, *G-functions and geometry*, Aspects of Mathematics, Vol. E13, Vieweg, Braunschweig, 1989.

[2] ——, *Mumford-Tate groups of mixed Hodge structures and the theorem of the fixed part*, Compositio Math. 82 (1992), pp. 1–24.

[3] S. ARAKELOV, *Families of algebraic curves with fixed degeneracies*, Math. USSR Izv. 5 (1971), pp. 1277–1302.

[4] I.N. BERNSTEIN, I.M. GEL'FAND and S.I. GEL'FAND, *Schubert cells and cohomoloy of the spaces G/P*, Russian Math. Surv. 28 (1973), Nr. 3, pp. 3–26.

[5] R. BISHOP and R. CRITTENDEN, *Geometry of manifolds*, Academic Press, New York and London, 1964.

[6] A. BOREL, *Introduction aux groupes arithmétiques*, Actualités scientifiques et industrielles, Vol. 1341, Hermann, Paris, 1969.

[7] ——, *Linear algebraic groups (second edition)*, Graduate texts in mathematics, Vol. 126, Springer-verlag, New York, 1991.

[8] M. BOROVOI, *The Hodge group and the algebra of endomorphisms of an abelian variety*, in Questions of group theory and homological algebra, A. Onishchik, ed., Yaroslav, 1981. (In Russian.)

[9] J-L. BRYLINSKI and S. ZUCKER, *An overview of recent advances in Hodge theory*, in Several complex variables VI, Encyclopaedia of mathematical sciences, Vol. 69, Springer-Verlag, Berlin, 1990, pp. 39–142.

[10] E. CATTANI, P. DELIGNE, and A. KAPLAN, *On the locus of Hodge classes*, J. A.M.S. 8 (1995), pp. 483–506.

[11] C. CHEVALLEY, *Théorie des groupes de Lie, Tome II: Groupes Algébriques*, Actualités scientifiques et industrielles, Vol. 1152, Hermann & C$^{\text{ie}}$, Paris, 1952.

[12] R. COLEMAN, *Torsion points on curves*, in Galois representations and arithmetic algebraic geometry, Y. Ihara, ed., Adv. Studies in Pure Math., Vol. 12, North-Holland, Amsterdam, 1987, pp. 235–247.

[13] A. J. DE JONG and R. NOOT, *Jacobians with complex multiplication*, in Arithmetic Algebraic Geometry, G. van der Geer, F. Oort and J. Steenbrink, eds., Progress in Math., Vol. 89, Birkhäuser, Boston, 1991, pp. 177–192.

[14] P. DELIGNE, *Travaux de Shimura*, in Séminaire Bourbaki, Exposé 389, Février 1971, pp. 123–165.

[15] ——, *La conjecture de Weil pour les surfaces K3*, Invent. Math. 15 (1972), pp. 206–226.

[16] ——, *Théorie de Hodge II*, Publ. Math. de l'I.H.E.S. 40 (1972), pp. 5–57.

[17] ——, *Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques*, in Automorphic Forms, Representations, and L-functions, Part 2, A. Borel and W. Casselman, eds., Proc. of Symp. in Pure Math., Vol. XXXIII, American Mathematical Society, 1979, pp. 247–290.

[18] ——, *Hodge cycles on abelian varieties (Notes by J.S. Milne)*, in Hodge Cycles, Motives, and Shimura Varieties, Lecture Notes in Mathematics 900, Springer-Verlag, Berlin, 1982, pp. 9–100.

[19] G. FALTINGS, *Arakelov's theorem for abelian varieties*, Invent. Math. 73 (1983), pp. 337–347.

[20] S. HELGASON, *Differential Geometry, Lie groups, and symmetric spaces*, Academic Press, Inc., Orlando, 1978.

[21] N. M. KATZ, *Serre-Tate local moduli*, Exposé V$^{\text{bis}}$ in Surfaces algébriques, J. Giraud, L. Illusie, and M. Raynaud, eds., Lecture Notes in Mathematics 868, Springer-Verlag, Berlin, 1981, pp. 138–202.

[22] J. S. Milne, *Canonical models of (mixed) Shimura varieties and automorphic vector bundles*, in Automorphic forms, Shimura varieties, and L-functions, L. Clozel and J. S. Milne, eds., Persp. in Math., Vol. 10(I), Academic Press, Inc., 1990, pp. 283–414.

[23] ——, *Shimura varieties and motives*, in Motives, Part 2, U. Jannsen, S. Kleiman, and J-P. Serre, eds., Proc. of Symp. in Pure Math., Vol. 55, American Mathematical Society, 1994, pp. 447–523.

[24] F. Oort, *Canonical liftings and dense sets of CM-points*, in: Arithmetic geometry, Proc. Cortona symposium 1994, F. Catanese, ed., Symposia Math., Vol. XXXVII, Cambridge University Press, 1997, pp. 228–234.

[25] C. Peters, *Rigidity for variations of Hodge structure and Arakelov-type finiteness theorems*, Compositio Math. 75 (1990), pp. 113–126.

[26] M.-H. Saito, *Classification of non-rigid families of abelian varieties*, Tôhoku Math. J. 45 (1993), pp. 159–189.

Address: Ben Moonen, Westfälische Wilhelms-Universität Münster, Mathematisches Institut, Einsteinstraße 62, 48149 Münster, Germany. *E-mail:* `moonen@math.uni-muenster.de`