Clay Mathematics Proceedings
Volume 4, 2005

Introduction to Shimura Varieties

J.S. Milne

Abstract. This is an introduction to the theory of Shimura varieties, or, in other words, to the arithmetic theory of automorphic functions and holomorphic automorphic forms.

Contents

Introduction

The arithmetic properties of elliptic modular functions and forms were extensively studied in the 1800s, culminating in the beautiful Kronecker Jugendtraum. Hilbert emphasized the importance of extending this theory to functions of several variables in the twelfth of his famous problems at the International Congress in 1900. The first tentative steps in this direction were taken by Hilbert himself and

© 2005 J. S. Milne

his students Blumenthal and Hecke in their study of what are now called Hilbert (or Hilbert-Blumenthal) modular varieties. As the theory of complex functions of several variables matured, other quotients of bounded symmetric domains by arithmetic groups were studied (Siegel, Braun, and others). However, the modern theory of Shimura varieties[1] only really began with the development of the theory of abelian varieties with complex multiplication by Shimura, Taniyama, and Weil in the mid-1950s, and with the subsequent proof by Shimura of the existence of canonical models for certain families of Shimura varieties. In two fundamental articles, Deligne recast the theory in the language of abstract reductive groups and extended Shimura's results on canonical models. Langlands made Shimura varieties a central part of his program, both as a source of representations of galois groups and as tests for the conjecture that all motivic L-functions are automorphic. These notes are an introduction to the theory of Shimura varieties from the point of view of Deligne and Langlands. Because of their brevity, many proofs have been omitted or only sketched.

Notations and conventions. Unless indicated otherwise, vector spaces are assumed to be finite dimensional and free $\mathbb{Z}$-modules are assumed to be of finite rank. The linear dual $\mathrm{Hom}(V,k)$ of a vector space (or module) V is denoted $V^\vee$. For a k-vector space V and a k-algebra R, $V(R)$ denotes $R\otimes_k V$ (and similarly for $\mathbb{Z}$-modules). By a lattice in an $\mathbb{R}$-vector space V, I mean a full lattice, i.e., a $\mathbb{Z}$-submodule generated by a basis for V. The algebraic closure of a field k is denoted k^{al}.

A superscript $^+$ (resp. $^\circ$) denotes a connected component relative to a real topology (resp. a zariski topology). For an algebraic group, we take the identity connected component. For example, $(O_n)^\circ = \mathrm{SO}_n$, $(\mathrm{GL}_n)^\circ = \mathrm{GL}_n$, and $\mathrm{GL}_n(\mathbb{R})^+$ consists of the $n\times n$ matrices with $\det > 0$. For an algebraic group G over $\mathbb{Q}$, $G(\mathbb{Q})^+ = G(\mathbb{Q})\cap G(\mathbb{R})^+$. Following Bourbaki, I require compact topological spaces to be separated.

Semisimple and reductive groups, whether algebraic or Lie, are required to be connected. A simple algebraic or Lie group is a semisimple group with no connected proper normal subgroups other than 1 (some authors say almost-simple). For a torus T, $X^*(T)$ denotes the character group of T. The inner automorphism defined by an element g is denoted $\mathrm{ad}(g)$. The derived group of a reductive group G is denoted G^{der} (it is a semisimple group). For more notations concerning reductive groups, see p303. For a finite extension of fields $L\supset F$ of characteristic zero, the torus over F obtained by restriction of scalars from $\mathbb{G}_m$ over L is denoted $(\mathbb{G}_m)_{L/F}$.

Throughout, I use the notations standard in algebraic geometry, which sometimes conflict with those used in other areas. For example, if G and G' are algebraic groups over a field k, then a homomorphism $G\to G'$ means a homomorphism defined over k; if K is a field containing k, then G_K is the algebraic group over K obtained by extension of the base field and $G(K)$ is the group of points of G with coordinates in K. If $\sigma\colon k\hookrightarrow K$ is a homomorphism of fields and V is an algebraic variety (or other algebro-geometric object) over k, then σV has its only possible meaning: apply σ to the coefficients of the equations defining V.

Let A and B be sets and let $\sim$ be an equivalence relation on A. If there exists a canonical surjection $A\to B$ whose fibres are the equivalence classes, then I say

[1]The term "Shimura variety" was introduced by Langlands (1976, 1977), although earlier "Shimura curve" had been used for the varieties of dimension one (Ihara 1968).

that B classifies the elements of A modulo $\sim$ or that it classifies the $\sim$-classes of elements of A.

A functor $F\colon \mathsf{A} \to \mathsf{B}$ is fully faithful if the maps $\mathrm{Hom}_{\mathsf{A}}(a, a') \to \mathrm{Hom}_{\mathsf{B}}(Fa, Fa')$ are bijective. The essential image of such a functor is the full subcategory of B whose objects are isomorphic to an object of the form Fa. Thus, a fully faithful functor $F\colon \mathsf{A} \to \mathsf{B}$ is an equivalence if and only if its essential image is B (Mac Lane 1998, p93).

References. In addition to those listed at the end, I refer to the following of my course notes (available at www.jmilne.org/math/).

AG: Algebraic Geometry, v5.0, February 20, 2005.
ANT: Algebraic Number Theory, v2.1, August 31, 1998.
CFT: Class Field Theory, v3.1, May 6, 1997.
FT: Fields and galois Theory, v4.0, February 19, 2005.
MF: Modular Functions and Modular Forms, v1.1, May 22, 1997.

Prerequisites. Beyond the mathematics that students usually acquire by the end of their first year of graduate work (a little complex analysis, topology, algebra, differential geometry,...), I assume familiarity with some algebraic number theory, algebraic geometry, algebraic groups, and elliptic modular curves.

Acknowledgements. I thank the Clay Mathematical Institute and the organizers for giving me the opportunity to lecture on Shimura varieties, the Fields Institute for providing a excellent setting for the Summer School, and the audience for its lively participation. Also, I thank Lizhen Ji and Gopal Prasad for their help with references, and F. Hörmann and others for alerting me to errors in earlier versions.

1. Hermitian symmetric domains

In this section, I describe the complex manifolds that play the role in higher dimensions of the complex upper half plane, or, equivalently, the open unit disk:

$$\{z \in \mathbb{C} \mid \Im(z) > 0\} = \mathcal{H}_1 \underset{-i\frac{z+1}{z-1} \leftarrow z}{\overset{z \to \frac{z-i}{z+i}}{\rightleftarrows}} \mathcal{D}_1 = \{z \in \mathbb{C} \mid |z| < 1\}.$$

This is a large topic, and I can do little more than list the definitions and results that we shall need.

Brief review of real manifolds. A ***manifold*** M ***of dimension*** n is a separated topological space that is locally isomorphic to an open subset of $\mathbb{R}^n$ and admits a countable basis of open subsets. A homeomorphism from an open subset of M onto an open subset of $\mathbb{R}^n$ is called a ***chart*** of M.

Smooth manifolds. I use smooth to mean C^∞. A ***smooth manifold*** is a manifold M endowed with a ***smooth structure***, i.e., a sheaf $\mathcal{O}_M$ of $\mathbb{R}$-valued functions such that $(M, \mathcal{O}_M)$ is locally isomorphic to $\mathbb{R}^n$ endowed with its sheaf of smooth functions. For an open $U \subset M$, the $f \in \mathcal{O}_M(U)$ are called the ***smooth functions*** on U. A smooth structure on a manifold M can be defined by a family $u_\alpha\colon U_\alpha \to \mathbb{R}^n$ of charts such that $M = \bigcup U_\alpha$ and the maps

$$u_\alpha \circ u_\beta^{-1}\colon u_\beta(U_\alpha \cap U_\beta) \to u_\alpha(U_\alpha \cap U_\beta)$$

are smooth for all α, β. A continuous map $\alpha\colon M \to N$ of smooth manifolds is ***smooth*** if it is a map of ringed spaces, i.e., f smooth on an open $V \subset N$ implies $f \circ \alpha$ smooth on $\alpha^{-1}(V)$.

Let $(M, \mathcal{O}_M)$ be a smooth manifold, and let $\mathcal{O}_{M,p}$ be the ring of germs of smooth functions at p. The ***tangent space*** T_pM to M at p is the $\mathbb{R}$-vector space of $\mathbb{R}$-derivations

$X_p\colon \mathcal{O}_{M,p} \to \mathbb{R}$. If $x^1, \ldots, x^n$ are local coordinates at p, then $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$ is a basis for T_pM and $dx^1, \ldots, dx^n$ is the dual basis.

Let U be an open subset of a smooth manifold M. A ***smooth vector field*** X on U is a family of tangent vectors $X_p \in T_p(M)$ indexed by $p \in U$, such that, for any smooth function f on an open subset of U, $p \mapsto X_pf$ is smooth. A ***smooth*** r***-tensor field*** on U is a family $t = (t_p)_{p\in M}$ of multilinear mappings $t_p\colon T_pM \times \cdots \times T_pM \to \mathbb{R}$ (r copies of T_pM) such that, for any smooth vector fields $X_1, \ldots, X_r$ on an open subset of U, $p \mapsto t_p(X_1, \ldots, X_r)$ is a smooth function. A ***smooth*** (r,s)***-tensor field*** is a family $t_p\colon (T_pM)^r \times (T_pM)^{\vee s} \to \mathbb{R}$ satisfying a similar condition. Note that to give a smooth $(1,1)$-field amounts to giving a family of endomorphisms $t_p\colon T_pM \to T_pM$ with the property that $p \mapsto t_p(X_p)$ is a smooth vector field for any smooth vector field X.

A ***riemannian manifold*** is a smooth manifold endowed with a ***riemannian metric***, i.e., a smooth 2-tensor field g such that, for all $p \in M$, g_p is symmetric and positive definite. In terms of local coordinates $x^1, \ldots, x^n$ at p,

$$g_p = \textstyle\sum g_{i,j}(p)dx^i \otimes dx^j, \text{ i.e., } g_p\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = g_{ij}(p).$$

A morphism of riemannian manifolds is called an ***isometry***.

A ***real Lie group***[2] G is a smooth manifold endowed with a group structure defined by smooth maps $g_1, g_2 \mapsto g_1g_2$, $g \mapsto g^{-1}$.

Brief review of hermitian forms. To give a complex vector space amounts to giving a real vector space V together with an endomorphism $J\colon V \to V$ such that $J^2 = -1$. A ***hermitian form*** on (V,J) is an $\mathbb{R}$-bilinear mapping $(\,|\,)\colon V \times V \to \mathbb{C}$ such that $(Ju|v) = i(u|v)$ and $(v|u) = \overline{(u|v)}$. When we write

$$(u|v) = \varphi(u,v) - i\psi(u,v), \quad \varphi(u,v),\ \psi(u,v) \in \mathbb{R}, \tag{1}$$

then φ and ψ are $\mathbb{R}$-bilinear, and

$$\varphi \text{ is symmetric} \qquad \varphi(Ju,Jv) = \varphi(u,v), \tag{2}$$

$$\psi \text{ is alternating} \qquad \psi(Ju,Jv) = \psi(u,v), \tag{3}$$

$$\psi(u,v) = -\varphi(u,Jv), \qquad \varphi(u,v) = \psi(u,Jv). \tag{4}$$

As $(u|u) = \varphi(u,u)$, $(\,|\,)$ is positive definite if and only if φ is positive definite. Conversely, if φ satisfies (2) (resp. ψ satisfies (3)), then the formulas (4) and (1) define a hermitian form:

$$(u|v) = \varphi(u,v) + i\varphi(u,Jv) \quad (\text{resp. } (u|v) = \psi(u,Jv) - i\psi(u,v)) \tag{5}$$

Complex manifolds. A $\mathbb{C}$-valued function on an open subset U of $\mathbb{C}^n$ is ***analytic*** if it admits a power series expansion in a neighbourhod of each point of U. A ***complex manifold*** is a manifold M endowed with a ***complex structure***, i.e., a sheaf $\mathcal{O}_M$ of $\mathbb{C}$-valued functions such that $(M, \mathcal{O}_M)$ is locally isomorphic to $\mathbb{C}^n$ with its sheaf of analytic functions. A complex structure on a manifold M can be defined by a family $u_\alpha\colon U_\alpha \to \mathbb{C}^n$ of charts such that $M = \bigcup U_\alpha$ and the maps $u_\alpha \circ u_\beta^{-1}$ are analytic for all α, β. Such a family also makes M into a smooth manifold denoted M^∞. A continuous map $\alpha\colon M \to N$ of complex manifolds is ***analytic*** if it is a map of ringed spaces. A ***riemann surface*** is a one-dimensional complex manifold.

A ***tangent vector*** at a point p of a complex manifold is a $\mathbb{C}$-derivation $\mathcal{O}_{M,p} \to \mathbb{C}$. The tangent spaces T_pM (M as a complex manifold) and T_pM^∞ (M as a smooth manifold) can be identified. Explicitly, complex local coordinates $z^1, \ldots, z^n$ at a point p of M define real local coordinates $x^1, \ldots, x^n, y^1, \ldots, y^n$ with $z^r = x^r + iy^r$.

[2]According to a theorem of Lie, this is equivalent to the usual definition in which "smooth" is replaced by "real-analytic".

The real and complex tangent spaces have bases $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial y^1}, \ldots, \frac{\partial}{\partial y^n}$ and $\frac{\partial}{\partial z^1}, \ldots, \frac{\partial}{\partial z^n}$ respectively. Under the natural identification of the two spaces, $\frac{\partial}{\partial z^r} = \frac{1}{2}\left(\frac{\partial}{\partial x^r} - i\frac{\partial}{\partial y^r}\right)$.

A $\mathbb{C}$-valued function f on an open subset U of $\mathbb{C}^n$ is ***holomorphic*** if it is holomorphic (i.e., differentiable) separately in each variable. As in the one-variable case, f is holomorphic if and only if it is analytic (Hartog's theorem, Taylor 2002, 2.2.3), and so we can use the terms interchangeably.

Recall that a $\mathbb{C}$-valued function f on $U \subset \mathbb{C}$ is holomorphic if and only if it is smooth (as a function of two real variables) and satisfies the Cauchy-Riemann condition. This condition has a geometric interpretation: it requires that $df_p \colon T_pU \to T_{f(p)}\mathbb{C}$ be $\mathbb{C}$-linear for all $p \in U$. It follows that a smooth $\mathbb{C}$-valued function f on $U \subset \mathbb{C}^n$ is holomorphic if and only if the maps $df_p \colon T_pU \to T_{f(p)}\mathbb{C}$ are $\mathbb{C}$-linear for all $p \in U$.

An ***almost-complex structure*** on a smooth manifold M is a smooth tensor field $(J_p)_{p\in M}$, $J_p \colon T_pM \to T_pM$, such that $J_p^2 = -1$ for all p, i.e., it is a smoothly varying family of complex structures on the tangent spaces. A complex structure on a smooth manifold endows it with an almost-complex structure. In terms of complex local coordinates $z^1, \ldots, z^n$ in a neighbourhood of a point p on a complex manifold and the corresponding real local coordinates $x^1, \ldots, y^n$, J_p acts by

$$\frac{\partial}{\partial x^r} \mapsto \frac{\partial}{\partial y^r}, \quad \frac{\partial}{\partial y^r} \mapsto -\frac{\partial}{\partial x^r}. \tag{6}$$

It follows from the last paragraph that the functor from complex manifolds to almost-complex manifolds is fully faithful: a smooth map $\alpha \colon M \to N$ of complex manifolds is holomorphic (analytic) if the maps $d\alpha_p \colon T_pM \to T_{\alpha(p)}N$ are $\mathbb{C}$-linear for all $p \in M$. Not every almost-complex structure on a smooth manifold arises from a complex structure — those that do are said to be ***integrable***. An almost-complex structure J on a smooth manifold is integrable if M can be covered by charts on which J takes the form (6) (because this condition forces the transition maps to be holomorphic).

A ***hermitian metric*** on a complex (or almost-complex) manifold M is a riemannian metric g such that

$$g(JX, JY) = g(X, Y) \text{ for all vector fields } X, Y. \tag{7}$$

According to (5), for each $p \in M$, g_p is the real part of a unique hermitian form h_p on T_pM, which explains the name. A ***hermitian manifold*** (M, g) is a complex manifold with a hermitian metric, or, in other words, it is a riemannian manifold with a complex structure such that J acts by isometries.

Hermitian symmetric spaces. A manifold (riemannian, hermitian, ...) is said to be ***homogeneous*** if its automorphism group acts transitively. It is ***symmetric*** if, in addition, at some point p there is an involution s_p (the ***symmetry at*** p) having p as an isolated fixed point. This means that s_p is an automorphism such that $s_p^2 = 1$ and that p is the only fixed point of s_p in some neighbourhood of p.

For a riemannian manifold (M, g), the automorphism group is the group $\mathrm{Is}(M, g)$ of isometries. A connected symmetric riemannian manifold is called a ***symmetric space***. For example, $\mathbb{R}^n$ with the standard metric $g_p = \sum dx^i dx^i$ is a symmetric space — the translations are isometries, and $\mathbf{x} \mapsto -\mathbf{x}$ is a symmetry at 0.

For a hermitian manifold (M, g), the automorphism group is the group $\mathrm{Is}(M, g)$ of holomorphic isometries:

$$\mathrm{Is}(M, g) = \mathrm{Is}(M^\infty, g) \cap \mathrm{Hol}(M) \tag{8}$$

(intersection inside $\mathrm{Aut}(M^\infty)$; $\mathrm{Hol}(M)$ is the group of automorphisms of M as a complex manifold). A connected symmetric hermitian manifold is called a ***hermitian symmetric space***.

EXAMPLE 1.1. (a) The complex upper half plane $\mathcal{H}_1$ becomes a hermitian symmetric space when endowed with the metric $\frac{dxdy}{y^2}$. The action

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}), \quad z \in \mathcal{H}_1,$$

identifies $\mathrm{SL}_2(\mathbb{R})/\{\pm I\}$ with the group of holomorphic automorphisms of $\mathcal{H}_1$. For any $x+iy \in \mathcal{H}_1$, $x+iy = \left(\begin{smallmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{smallmatrix}\right) i$, and so $\mathcal{H}_1$ is homogeneous. The isomorphism $z \mapsto -1/z$ is a symmetry at $i \in \mathcal{H}_1$, and the riemannian metric $\frac{dxdy}{y^2}$ is invariant under the action of $\mathrm{SL}_2(\mathbb{R})$ and has the hermitian property (7).

(b) The projective line $\mathbb{P}^1(\mathbb{C})$ (= riemann sphere) becomes a hermitian symmetric space when endowed with the restriction (to the sphere) of the standard metric on $\mathbb{R}^3$. The group of rotations is transitive, and reflection along a geodesic (great circle) through a point is a symmetry. Both of these transformations leave the metric invariant.

(c) Any quotient $\mathbb{C}/\Lambda$ of $\mathbb{C}$ by a discrete additive subgroup Λ becomes a hermitian symmetric space when endowed with the standard metric. The group of translations is transitive, and $z \mapsto -z$ is a symmetry at 0.

Curvature. Recall that, for a plane curve, the curvature at a point p is $1/r$ where r is the radius of the circle that best approximates the curve at p. For a surface in 3-space, the principal curvatures at a point p are the maximum and minimum of the signed curvatures of the curves obtained by cutting the surface with planes through a normal at p (the sign is positive or negative according as the curve bends towards the normal or away). Although the principal curvatures depend on the embedding of the surface into $\mathbb{R}^3$, their product, the ***sectional curvature*** at p, does not (Gauss's Theorema Egregium) and so it is well-defined for any two-dimensional riemannian manifold. More generally, for a point p on any riemannian manifold M, one can define the ***sectional curvature*** $K(p, E)$ of the submanifold cut out by the geodesics tangent to a two-dimensional subspace E of T_pM. Intuitively, positive curvature means that the geodesics through a point converge, and negative curvature means that they diverge. The geodesics in the upper half plane are the half-lines and semicircles orthogonal to the real axis. Clearly, they diverge — in fact, this is Poincaré's famous model of noneuclidean geometry in which there are infinitely many "lines" through a point parallel to any fixed "line" not containing it. More prosaically, one can compute that the sectional curvature is -1. The Gauss curvature of $\mathbb{P}^1(\mathbb{C})$ is obviously positive, and that of $\mathbb{C}/\Lambda$ is zero.

The three types of hermitian symmetric spaces. The group of isometries of a symmetric space (M, g) has a natural structure of a Lie group (Helgason 1978, IV 3.2). For a hermitian symmetric space (M, g), the group $\mathrm{Is}(M, g)$ of holomorphic isometries is closed in the group of isometries of (M^∞, g) and so is also a Lie group.

There are three families of hermitian symmetric spaces (ibid, VIII; Wolf 1984, 8.7):

Name	example	simply connected?	curvature	$\mathrm{Is}(M,g)^+$
noncompact type	$\mathcal{H}_1$	yes	negative	adjoint, noncompact
compact type	$\mathbb{P}^1(\mathbb{C})$	yes	positive	adjoint, compact
euclidean	$\mathbb{C}/\Lambda$	not necessarily	zero	

A Lie group is ***adjoint*** if it is semisimple with trivial centre.

Every hermitian symmetric space, when viewed as hermitian manifold, decomposes into a product $M^0 \times M^- \times M^+$ with M^0 euclidean, M^- of noncompact type, and M^+ of compact type. The euclidean spaces are quotients of a complex space $\mathbb{C}^g$ by a discrete subgroup of translations. A hermitian symmetric space is ***irreducible*** if it is not the product of two hermitian symmetric spaces of lower dimension. Each of M^- and M^+ is a product of irreducible hermitian symmetric spaces, each of which has a simple isometry group.

We shall be especially interested in the hermitian symmetric spaces of noncompact type — they are called ***hermitian symmetric domains.***

EXAMPLE 1.2 (Siegel upper half space). The ***Siegel upper half space*** $\mathcal{H}_g$ ***of degree*** g consists of the symmetric complex $g \times g$ matrices with positive definite imaginary part, i.e.,

$$\mathcal{H}_g = \{Z = X + iY \in M_g(\mathbb{C}) \mid X = X^t, \quad Y > 0\}.$$

Note that the map $Z = (z_{ij}) \mapsto (z_{ij})_{j \geq i}$ identifies $\mathcal{H}_g$ with an open subset of $\mathbb{C}^{g(g+1)/2}$. The symplectic group $\mathrm{Sp}_{2g}(\mathbb{R})$ is the group fixing the alternating form $\sum_{i=1}^{g} x_i y_{-i} - \sum_{i=1}^{g} x_{-i} y_i$:

$$\mathrm{Sp}_{2g}(\mathbb{R}) = \left\{ \left(\begin{array}{cc} A & B \\ C & D \end{array} \right) \middle| \begin{array}{ll} A^tC = C^tA & A^tD - C^tB = I_g \\ D^tA - B^tC = I_g & B^tD = D^tB \end{array} \right\}.$$

The group $\mathrm{Sp}_{2g}(\mathbb{R})$ acts transitively on $\mathcal{H}_g$ by

$$\left(\begin{array}{cc} A & B \\ C & D \end{array} \right) Z = (AZ + B)(CZ + D)^{-1}.$$

The matrix $\left(\begin{smallmatrix} 0 & -I_g \\ I_g & 0 \end{smallmatrix} \right)$ acts as an involution on $\mathcal{H}_g$, and has iI_g as its only fixed point. Thus, $\mathcal{H}_g$ is homogeneous and symmetric as a complex manifold, and we shall see in (1.4) below that $\mathcal{H}_g$ is in fact a hermitian symmetric domain.

Example: Bounded symmetric domains. A ***domain*** D in $\mathbb{C}^n$ is a nonempty open connected subset. It is ***symmetric*** if the group $\mathrm{Hol}(D)$ of holomorphic automorphisms of D (as a complex manifold) acts transitively and for some point there exists a holomorphic symmetry. For example, $\mathcal{H}_1$ is a symmetric domain and $\mathcal{D}_1$ is a bounded symmetric domain.

THEOREM 1.3. *Every bounded domain has a canonical hermitian metric (called the Bergman(n) metric). Moreover, this metric has negative curvature.*

PROOF (SKETCH): Initially, let D be any domain in $\mathbb{C}^n$. The holomorphic square-integrable functions $f\colon D \to \mathbb{C}$ form a Hilbert space $H(D)$ with inner product $(f|g) = \int_D f\bar{g}dv$. There is a unique (Bergman kernel) function $K\colon D \times D \to \mathbb{C}$ such that

(a) the function $z \mapsto K(z,\zeta)$ lies in $H(D)$ for each ζ,
(b) $K(z,\zeta) = \overline{K(\zeta,z)}$, and
(c) $f(z) = \int K(z,\zeta)f(\zeta)dv(\zeta)$ for all $f \in H(D)$.

For example, for any complete orthonormal set $(e_m)_{m\in\mathbb{N}}$ in $H(D)$, $K(z,\zeta) = \sum_m e_m(z) \cdot \overline{e_m(\zeta)}$ is such a function. If D is bounded, then all polynomial functions on D are square-integrable, and so certainly $K(z,z) > 0$ for all z. Moreover, $\log(K(z,z))$ is smooth and the equations

$$h = \sum h_{ij} dz^i d\bar{z}^j, \quad h_{ij}(z) = \frac{\partial^2}{\partial z^i \partial \bar{z}^j} \log K(z,z),$$

define a hermitian metric on D, which can be shown to have negative curvature (Helgason 1978, VIII 3.3, 7.1; Krantz 1982, 1.4). □

The Bergman metric, being truly canonical, is invariant under the action $\mathrm{Hol}(D)$. Hence, a bounded symmetric domain becomes a hermitian symmetric domain for the Bergman metric. Conversely, it is known that every hermitian symmetric domain can be embedded into some $\mathbb{C}^n$ as a bounded symmetric domain. Therefore, a hermitian symmetric domain D has a unique hermitian metric that maps to the Bergman metric under every isomorphism of D with a bounded symmetric domain. On each irreducible factor, it is a multiple of the original metric.

Example 1.4. Let $\mathcal{D}_g$ be the set of symmetric complex matrices such that $I_g - \overline{Z}^t Z$ is positive definite. Note that $(z_{ij}) \mapsto (z_{ij})_{j\geq i}$ identifies $\mathcal{D}_g$ as a bounded domain in $\mathbb{C}^{g(g+1)/2}$. The map $Z \mapsto (Z - iI_g)(Z + iI_g)^{-1}$ is an isomorphism of $\mathcal{H}_g$ onto $\mathcal{D}_g$. Therefore, $\mathcal{D}_g$ is symmetric and $\mathcal{H}_g$ has an invariant hermitian metric: they are both hermitian symmetric domains.

Automorphisms of a hermitian symmetric domain.

Lemma 1.5. *Let (M,g) be a symmetric space, and let $p \in M$. Then the subgroup K_p of $\mathrm{Is}(M,g)^+$ fixing p is compact, and*

$$a \cdot K_p \mapsto a \cdot p \colon \mathrm{Is}(M,g)^+/K_p \to M$$

is an isomorphism of smooth manifolds. In particular, $\mathrm{Is}(M,g)^+$ acts transitively on M.

Proof. For any riemannian manifold (M,g), the compact-open topology makes $\mathrm{Is}(M,g)$ into a locally compact group for which the stabilizer K'_p of a point p is compact (Helgason 1978, IV 2.5). The Lie group structure on $\mathrm{Is}(M,g)$ noted above is the unique such structure compatible with the compact-open topology (ibid. II 2.6). An elementary argument (e.g., MF 1.2) now shows that $\mathrm{Is}(M,g)/K'_p \to M$ is a homeomorphism, and it follows that the map $a \mapsto ap\colon \mathrm{Is}(M,g) \to M$ is open. Write $\mathrm{Is}(M,g)$ as a finite disjoint union $\mathrm{Is}(M,g) = \bigsqcup_i \mathrm{Is}(M,g)^+ a_i$ of cosets of $\mathrm{Is}(M,g)^+$. For any two cosets the open sets $\mathrm{Is}(M,g)^+ a_i p$ and $\mathrm{Is}(M,g)^+ a_j p$ are either disjoint or equal, but, as M is connected, they must all be equal, which shows that $\mathrm{Is}(M,g)^+$ acts transitively. Now $\mathrm{Is}(M,g)^+/K_p \to M$ is a homeomorphism, and it follows that it is a diffeomorphism (Helgason 1978, II 4.3a). □

Proposition 1.6. *Let (M,g) be a hermitian symmetric domain. The inclusions*

$$\mathrm{Is}(M^\infty, g) \supset \mathrm{Is}(M,g) \subset \mathrm{Hol}(M)$$

give equalities:

$$\mathrm{Is}(M^{\infty}, g)^{+} = \mathrm{Is}(M, g)^{+} = \mathrm{Hol}(M)^{+}.$$

Therefore, $\mathrm{Hol}(M)^{+}$ *acts transitively on* M*, and* $\mathrm{Hol}(M)^{+}/K_p \cong M^{\infty}$.

PROOF. The first equality is proved in Helgason 1978, VIII 4.3, and the second can be proved similarly. The rest of the statement follows from (1.5). □

Let H be a connected real Lie group. There need not be an algebraic group G over $\mathbb{R}$ such that[3] $G(\mathbb{R})^{+} = H$. However, if H has a faithful finite-dimensional representation $H \hookrightarrow \mathrm{GL}(V)$, then there exists an algebraic group $G \subset \mathrm{GL}(V)$ such that $\mathrm{Lie}(G) = [\mathfrak{h}, \mathfrak{h}]$ (inside $\mathfrak{gl}(V)$) where $\mathfrak{h} = \mathrm{Lie}(H)$ (Borel 1991, 7.9). If H, in addition, is semisimple, then $[\mathfrak{h}, \mathfrak{h}] = \mathfrak{h}$ and so $\mathrm{Lie}(G) = \mathfrak{h}$ and $G(\mathbb{R})^{+} = H$ (inside $\mathrm{GL}(V)$). This observation applies to any connected adjoint Lie group and, in particular, to $\mathrm{Hol}(M)^{+}$, because the adjoint representation on the Lie algebra is faithful.

PROPOSITION 1.7. *Let* (M, g) *be a hermitian symmetric domain, and let* $\mathfrak{h} = \mathrm{Lie}(\mathrm{Hol}(M)^{+})$. *There is a unique connected algebraic subgroup* G *of* $\mathrm{GL}(\mathfrak{h})$ *such that*

$$G(\mathbb{R})^{+} = \mathrm{Hol}(M)^{+} \qquad (\textit{inside } \mathrm{GL}(\mathfrak{h})).$$

For such a G*,*

$$G(\mathbb{R})^{+} = G(\mathbb{R}) \cap \mathrm{Hol}(M) \qquad (\textit{inside } \mathrm{GL}(\mathfrak{h}));$$

therefore $G(\mathbb{R})^{+}$ *is the stablizer in* $G(\mathbb{R})$ *of* M.

PROOF. The first statement was proved above, and the second follows from Satake 1980, 8.5. □

EXAMPLE 1.8. The map $z \mapsto \overline{z}^{-1}$ is an antiholomorphic isometry of $\mathcal{H}_1$, and every isometry of $\mathcal{H}_1$ is either holomorphic or differs from $z \mapsto \overline{z}^{-1}$ by a holomorphic isometry. In this case, $G = \mathrm{PGL}_2$, and $\mathrm{PGL}_2(\mathbb{R})$ acts holomorphically on $\mathbb{C} \smallsetminus \mathbb{R}$ with $\mathrm{PGL}_2(\mathbb{R})^{+}$ as the stabilizer of $\mathcal{H}_1$.

The homomorphism $u_p \colon U_1 \to \mathrm{Hol}(D)$**.** Let $U_1 = \{z \in \mathbb{C} \mid |z| = 1\}$ (the circle group).

THEOREM 1.9. *Let* D *be a hermitian symmetric domain. For each* $p \in D$*, there exists a unique homomorphism* $u_p \colon U_1 \to \mathrm{Hol}(D)$ *such that* $u_p(z)$ *fixes* p *and acts on* T_pD *as multiplication by* z.

EXAMPLE 1.10. Let $p = i \in \mathcal{H}_1$, and let $h \colon \mathbb{C}^{\times} \to \mathrm{SL}_2(\mathbb{R})$ be the homomorphism $z = a + ib \mapsto \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$. Then $h(z)$ acts on the tangent space $T_i\mathcal{H}_1$ as multiplication by $z/\overline{z}$, because $\frac{d}{dz}\left(\frac{az+b}{-bz+a}\right)|_i = \frac{a^2+b^2}{(a-bi)^2}$. For $z \in U_1$, choose a square root $\sqrt{z} \in U_1$, and set $u(z) = h(\sqrt{z}) \mod \pm I$. Then $u(z)$ is independent of the choice of $\sqrt{z}$ because $h(-1) = -I$. Therefore, u is a well-defined homomorphism $U_1 \to \mathrm{PSL}_2(\mathbb{R})$ such that $u(z)$ acts on the tangent space $T_i\mathcal{H}_1$ as multiplication by z.

Because of the importance of the theorem, I sketch a proof.

[3]For example, the (topological) fundamental group of $\mathrm{SL}_2(\mathbb{R})$ is $\mathbb{Z}$, and so $\mathrm{SL}_2(\mathbb{R})$ has many proper covering groups (even of finite degree). None of them is algebraic.

PROPOSITION 1.11. *Let (M, g) be symmetric space. The symmetry s_p at p acts as -1 on T_pM, and, for any geodesic γ with $\gamma(0) = p$, $s_p(\gamma(t)) = \gamma(-t)$. Moreover, (M, g) is (geodesically) complete.*

PROOF. Because $s_p^2 = 1$, $(ds_p)^2 = 1$, and so ds_p acts semisimply on T_pM with eigenvalues ± 1. Recall that for any tangent vector X at p, there is a unique geodesic $\gamma\colon I \to M$ with $\gamma(0) = p$, $\dot\gamma(0) = X$. If $(ds_p)(X) = X$, then $s_p \circ \gamma$ is a geodesic sharing these properties, and so p is not an isolated fixed point of s_p. This proves that only -1 occurs as an eigenvalue. If $(ds_p)(X) = -X$, then $s_p \circ \gamma$ and $t \mapsto \gamma(-t)$ are geodesics through p with velocity $-X$, and so are equal. For the final statement, see Boothby 1975, VII 8.4. □

By a ***canonical tensor*** on a symmetric space (M, g), I mean any tensor canonically derived from g, and hence fixed by any isometry of (M, g).

PROPOSITION 1.12. *On a symmetric space (M, g) every canonical r-tensor with r odd is zero. In particular, parallel translation of two-dimensional subspaces does not change the sectional curvature.*

PROOF. Let t be a canonical r-tensor. Then

$$t_p = t_p \circ (ds_p)^r \overset{1.11}{=} (-1)^r t_p,$$

and so $t = 0$ if r is odd. For the second statement, let ∇ be the riemannian connection, and let R be the corresponding curvature tensor (Boothby 1975, VII 3.2, 4.4). Then ∇R is an odd tensor, and so is zero. This implies that parallel translation of 2-dimensional subspaces does not change the sectional curvature. □

PROPOSITION 1.13. *Let (M, g) and (M', g') be riemannian manifolds in which parallel translation of 2-dimensional subspaces does not change the sectional curvature. Let $a\colon T_pM \to T_{p'}M'$ be a linear isometry such that $K(p, E) = K(p', aE)$ for every 2-dimensional subspace $E \subset T_pM$. Then $\exp_p(X) \mapsto \exp_{p'}(aX)$ is an isometry of a neighbourhood of p onto a neighbourhood of p'.*

PROOF. This follows from comparing the expansions of the riemann metrics in terms of normal geodesic coordinates. See Wolf 1984, 2.3.7. □

PROPOSITION 1.14. *If in (1.13) M and M' are complete, connected, and simply connected, then there is a unique isometry $\alpha\colon M \to M'$ such that $\alpha(p) = p'$ and $d\alpha_p = a$.*

PROOF. See Wolf 1984, 2.3.12. □

I now complete the sketch of the proof of Theorem 1.9. Each z with $|z| = 1$ defines an automorphism of (T_pD, g_p), and one checks that it preserves sectional curvatures. According to (1.11, 1.12, 1.14), there exists a unique isometry $u_p(z)\colon D \to D$ such that $du_p(z)_p$ is multiplication by z. It is holomorphic because it is $\mathbb{C}$-linear on the tangent spaces. The isometry $u_p(z) \circ u_p(z')$ fixes p and acts as multiplication by zz' on T_pD, and so equals $u_p(zz')$.

Cartan involutions. Let G be a connected algebraic group over $\mathbb{R}$, and let $g \mapsto \overline{g}$ denote complex conjugation on $G(\mathbb{C})$. An involution θ of G (as an algebraic group over $\mathbb{R}$) is said to be ***Cartan*** if the group

$$G^{(\theta)}(\mathbb{R}) \overset{\text{df}}{=} \{g \in G(\mathbb{C}) \mid g = \theta(\overline{g})\} \tag{9}$$

is compact.

EXAMPLE 1.15. Let $G = \mathrm{SL}_2$, and let $\theta = \mathrm{ad}\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. For $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{C})$, we have

$$\theta\left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \cdot \overline{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \cdot \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)^{-1} = \begin{pmatrix} \overline{d} & -\overline{c} \\ -\overline{b} & \overline{a} \end{pmatrix}.$$

Thus,

$$\begin{aligned} \mathrm{SL}_2^{(\theta)}(\mathbb{R}) &= \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{C}) \mid d = \overline{a},\ c = -\overline{b}\right\} \\ &= \left\{\begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C}) \mid |a|^2 + |b|^2 = 1\right\} = \mathrm{SU}_2, \end{aligned}$$

which is compact, being a closed bounded set in $\mathbb{C}^2$. Thus θ is a Cartan involution for SL_2.

THEOREM 1.16. *There exists a Cartan involution if and only if G is reductive, in which case any two are conjugate by an element of $G(\mathbb{R})$.*

PROOF. See Satake 1980, I 4.3. $\square$

EXAMPLE 1.17. Let G be a connected algebraic group over $\mathbb{R}$.

(a) The identity map on G is a Cartan involution if and only if $G(\mathbb{R})$ is compact.

(b) Let $G = \mathrm{GL}(V)$ with V a real vector space. The choice of a basis for V determines a transpose operator $M \mapsto M^t$, and $M \mapsto (M^t)^{-1}$ is obviously a Cartan involution. The theorem says that all Cartan involutions of G arise in this way.

(c) Let $G \hookrightarrow \mathrm{GL}(V)$ be a faithful representation of G. Then G is reductive if and only if G is stable under $g \mapsto g^t$ for a suitable choice of a basis for V, in which case the restriction of $g \mapsto (g^t)^{-1}$ to G is a Cartan involution; all Cartan involutions of G arise in this way from the choice of a basis for V (Satake 1980, I 4.4).

(d) Let θ be an involution of G. There is a unique real form $G^{(\theta)}$ of $G_{\mathbb{C}}$ such that complex conjugation on $G^{(\theta)}(\mathbb{C})$ is $g \mapsto \theta(\overline{g})$. Then, $G^{(\theta)}(\mathbb{R})$ satisfies (9), and we see that the Cartan involutions of G correspond to the compact forms of $G_{\mathbb{C}}$.

PROPOSITION 1.18. *Let G be a connected algebraic group over $\mathbb{R}$. If $G(\mathbb{R})$ is compact, then every finite-dimensional real representation of $G \to \mathrm{GL}(V)$ carries a G-invariant positive definite symmetric bilinear form; conversely, if one faithful finite-dimensional real representation of G carries such a form, then $G(\mathbb{R})$ is compact.*

PROOF. Let $\rho\colon G \to \mathrm{GL}(V)$ be a real representation of G. If $G(\mathbb{R})$ is compact, then its image H in $\mathrm{GL}(V)$ is compact. Let dh be the Haar measure on H, and choose a positive definite symmetric bilinear form $\langle\,|\,\rangle$ on V. Then the form

$$\langle u|v\rangle' = \int_H \langle hu|hv\rangle dh$$

is G-invariant, and it is still symmetric, positive definite, and bilinear. For the converse, choose an orthonormal basis for the form. Then $G(\mathbb{R})$ becomes identified with a closed set of real matrices A such that $A^t \cdot A = I$, which is bounded. $\square$

REMARK 1.19. The proposition can be restated for complex representations: if $G(\mathbb{R})$ is compact then every finite-dimensional complex representation of G carries a G-invariant positive definite Hermitian form; conversely, if some faithful finite-dimensional complex representation of G carries a G-invariant positive definite Hermitian form, then G is compact. (In this case, $G(\mathbb{R})$ is a subgroup of a unitary group instead of an orthogonal group. For a sesquilinear form φ to be G-invariant means that $\varphi(gu, \overline{g}v) = \varphi(u, v)$, $g \in G(\mathbb{C})$, $u, v \in V$.)

Let G be a real algebraic group, and let C be an element of $G(\mathbb{R})$ whose square is central (so that $\mathrm{ad}C$ is an involution). A ***C-polarization*** on a real representation V of G is a G-invariant bilinear form φ such that the form φ_C,

$$(u, v) \mapsto \varphi(u, Cv),$$

is symmetric and positive definite.

PROPOSITION 1.20. *If* $\mathrm{ad}C$ *is a Cartan involution of* G*, then every finite-dimensional real representation of* G *carries a* C*-polarization; conversely, if one faithful finite-dimensional real representation of* G *carries a* C*-polarization, then* $\mathrm{ad}C$ *is a Cartan involution.*

PROOF. An $\mathbb{R}$-bilinear form φ on a real vector space V defines a sesquilinear form φ' on $V(\mathbb{C})$,

$$\varphi' \colon V(\mathbb{C}) \times V(\mathbb{C}) \to \mathbb{C}, \quad \varphi'(u, v) = \varphi_{\mathbb{C}}(u, \overline{v}).$$

Moreover, φ' is hermitian (and positive definite) if and only if φ is symmetric (and positive definite).

Let $\rho \colon G \to \mathrm{GL}(V)$ be a real representation of G. For any G-invariant bilinear form φ on V, $\varphi_{\mathbb{C}}$ is $G(\mathbb{C})$-invariant, and so

$$\varphi'(gu, \overline{g}v) = \varphi'(u, v), \quad \text{all } g \in G(\mathbb{C}), \quad u, v \in V(\mathbb{C}). \tag{10}$$

On replacing v with Cv in this equality, we find that

$$\varphi'(gu, C(C^{-1}\overline{g}C)v) = \varphi'(u, Cv), \quad \text{all } g \in G(\mathbb{C}), \quad u, v \in V(\mathbb{C}), \tag{11}$$

which says that φ'_C is invariant under $G^{(\mathrm{ad}C)}$.

If ρ is faithful and φ is a C-polarization, then φ'_C is a positive definite hermitian form, and so $G^{(\mathrm{ad}C)}(\mathbb{R})$ is compact (1.19): $\mathrm{ad}C$ is a Cartan involution.

Conversely, if $G^{(\mathrm{ad}C)}(\mathbb{R})$ is compact, then every real representation $G \to \mathrm{GL}(V)$ carries a $G^{(\mathrm{ad}C)}(\mathbb{R})$-invariant positive definite symmetric bilinear form φ (1.18). Similar calculations to the above show that $\varphi_{C^{-1}}$ is a C-polarization on V. $\square$

Representations of U_1. Let T be a torus over a field k, and let K be a galois extension of k splitting T. To give a representation ρ of T on a k-vector space V amounts to giving an $X^*(T)$-grading $V(K) = \bigoplus_{\chi \in X^*(T)} V_\chi$ on $V(K) =_{\mathrm{df}} K \otimes_k V$ with the property that

$$\sigma(V_\chi) = V_{\sigma\chi}, \quad \text{all } \sigma \in \mathrm{Gal}(K/k), \quad \chi \in X^*(T).$$

Here V_χ is the subspace of $K \otimes_k V$ on which T acts through χ:

$$\rho(t)v = \chi(t) \cdot v, \quad \text{for } v \in V_\chi, \quad t \in T(K).$$

If $V_\chi \neq 0$, we say that χ ***occurs*** in V.

When we regard U_1 as a real algebraic torus, its characters are $z \mapsto z^n$, $n \in \mathbb{Z}$. Thus, $X^*(U_1) \cong \mathbb{Z}$, and complex conjugation acts on $X^*(U_1)$ as multiplication by -1. Therefore a representation of U_1 on a real vector space V corresponds to a grading $V(\mathbb{C}) = \oplus_{n \in \mathbb{Z}} V^n$ with the property that $V(\mathbb{C})^{-n} = \overline{V(\mathbb{C})^n}$ (complex conjugate). Here V^n is the subspace of $V(\mathbb{C})$ on which z acts as z^n. Note that

$V(\mathbb{C})^0 = \overline{V(\mathbb{C})^0}$ and so it is defined over $\mathbb{R}$, i.e., $V(\mathbb{C})^0 = V^0(\mathbb{C})$ for V^0 the subspace $V \cap V(\mathbb{C})^0$ of V (see AG 16.7). The natural map

$$V/V^0 \to V(\mathbb{C})/\textstyle\bigoplus_{n\leq 0} V(\mathbb{C})^n \cong \bigoplus_{n>0} V(\mathbb{C})^n \tag{12}$$

is an isomorphism. From this discussion, we see that every real representation of U_1 is a direct sum of representations of the following types:

(a) $V = \mathbb{R}$ with U_1 acting trivially (so $V(\mathbb{C}) = V^0$);
(b) $V = \mathbb{R}^2$ with $z = x + iy \in U_1(\mathbb{R})$ acting as $\left(\begin{smallmatrix} x & y \\ -y & x \end{smallmatrix}\right)^n$, $n > 0$ (so $V(\mathbb{C}) = V^n \oplus V^{-n}$).

Classification of hermitian symmetric domains in terms of real groups. The representations of U_1 have the same description whether we regard it as a Lie group or an algebraic group, and so every homomorphism $U_1 \to \mathrm{GL}(V)$ of Lie groups is algebraic. It follows that the homomorphism $u_p\colon U_1 \to \mathrm{Hol}(D)^+ \cong G(\mathbb{R})^+$ (see 1.9, 1.7) is algebraic.

THEOREM 1.21. *Let D be a hermitian symmetric domain, and let G be the associated real adjoint algebraic group (1.7). The homomorphism $u_p\colon U_1 \to G$ attached to a point p of D has the following properties:*

(a) *only the characters $z, 1, z^{-1}$ occur in the representation of U_1 on* $\mathrm{Lie}(G)_{\mathbb{C}}$ *defined by u_p;*
(b) $\mathrm{ad}(u_p(-1))$ *is a Cartan involution;*
(c) $u_p(-1)$ *does not project to* 1 *in any simple factor of G.*

Conversely, let G be a real adjoint algebraic group, and let $u\colon U_1 \to G$ satisfy (a), (b), and (c). Then the set D of conjugates of u by elements of $G(\mathbb{R})^+$ has a natural structure of a hermitian symmetric domain for which $G(\mathbb{R})^+ = \mathrm{Hol}(D)^+$ and $u(-1)$ is the symmetry at u (regarded as a point of D).

PROOF (SKETCH): Let D be a hermitian symmetric domain, and let G be the associated group (1.7). Then $G(\mathbb{R})^+/K_p \cong D$ where K_p is the group fixing p (see 1.6). For $z \in U_1$, $u_p(z)$ acts on the $\mathbb{R}$-vector space

$$\mathrm{Lie}(G)/\mathrm{Lie}(K_p) \cong T_pD$$

as multiplication by z, and it acts on $\mathrm{Lie}(K_p)$ trivially. From this, (a) follows.

The symmetry s_p at p and $u_p(-1)$ both fix p and act as -1 on T_pD (see 1.11); they are therefore equal (1.14). It is known that the symmetry at a point of a symmetric space gives a Cartan involution of G if and only if the space has negative curvature (see Helgason 1978, V 2; the real form of G defined by $\mathrm{ad}s_p$ is that attached to the compact dual of the symmetric space). Thus (b) holds.

Finally, if the projection of $u(-1)$ into a simple factor of G were trivial, then that factor would be compact (by (b); see 1.17a), and D would have an irreducible factor of compact type.

For the converse, let D be the set of $G(\mathbb{R})^+$-conjugates of u. The centralizer K_u of u in $G(\mathbb{R})^+$ is contained in $\{g \in G(\mathbb{C}) \mid g = u(-1) \cdot \overline{g} \cdot u(-1)^{-1}\}$, which, according to (b), is compact. As K_u is closed, it also is compact. The equality $D = (G(\mathbb{R})^+/K_u) \cdot u$ endows D with the structure of smooth (even real-analytic) manifold. For this structure, the tangent space to D at u,

$$T_uD = \mathrm{Lie}(G)/\mathrm{Lie}(K_u),$$

which, because of (a), can be identified with the subspace of $\mathrm{Lie}(G)_{\mathbb{C}}$ on which $u(z)$ acts as z (see (12)). This endows T_uD with a $\mathbb{C}$-vector space structure for which $u(z)$, $z \in U_1$, acts as multiplication by z. Because D is homogeneous, this gives it the structure of an almost-complex manifold, which can be shown to integrable (Wolf 1984, 8.7.9). The action of K_u on D defines an action of it on T_uD. Because K_u is compact, there is a K_u-invariant positive definite form on T_uD (see 1.18), and because $J = u(i) \in K_u$, any such form will have the hermitian property (7). Choose one, and use the homogeneity of D to move it to each tangent space. This will make D into a hermitian symmetric space, which will be a hermitian symmetric domain because each simple factor of its automorphism group is a noncompact semisimple group (because of (b,c)). □

COROLLARY 1.22. *There is a natural one-to-one correspondence between isomorphism classes of pointed hermitian symmetric domains and pairs (G, u) consisting of a real adjoint Lie group and a nontrivial homomorphism $u\colon U_1 \to G(\mathbb{R})$ satisfying (a), (b), (c).*

EXAMPLE 1.23. Let $u\colon U_1 \to \mathrm{PSL}_2(\mathbb{R})$ be as in (1.10). Then $u(-1) = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ and we saw in 1.15 that $\mathrm{ad}u(-1)$ is a Cartan involution of SL_2, hence also of PSL_2.

Classification of hermitian symmetric domains in terms of dynkin diagrams. Let G be a simple adjoint group over $\mathbb{R}$, and let u be a homomorphism $U_1 \to G$ satisfying (a) and (b) of Theorem 1.21. By base extension, we get an adjoint group $G_{\mathbb{C}}$, which is simple because it is an inner form of its compact form, and a cocharacter $\mu = u_{\mathbb{C}}$ of $G_{\mathbb{C}}$ satisfying the following condition:

> (*) in the action of $\mathbb{G}_m$ on $\mathrm{Lie}(G_{\mathbb{C}})$ defined by $\mathrm{ad} \circ \mu$, only the characters $z, 1, z^{-1}$ occur.

PROPOSITION 1.24. *The map $(G, u) \mapsto (G_{\mathbb{C}}, u_{\mathbb{C}})$ defines a bijection between the sets of isomorphism classes of pairs consisting of*

(a) *a simple adjoint group over $\mathbb{R}$ and a conjugacy class of $u\colon U_1 \to H$ satisfying (1.21a,b), and*
(b) *a simple adjoint group over $\mathbb{C}$ and a conjugacy class of cocharacters satisfying (*).*

PROOF. Let (G, μ) be as in (b), and let $g \mapsto \overline{g}$ denote complex conjugation on $G(\mathbb{C})$ relative to the unique compact real form of G (cf. 1.16). There is a real form H of G such that complex conjugation on $H(\mathbb{C}) = G(\mathbb{C})$ is $g \mapsto \mu(-1) \cdot \overline{g} \cdot \mu(-1)^{-1}$, and $u =_{\mathrm{df}} \mu|U_1$ takes values in $H(\mathbb{R})$. The pair (H, u) is as in (a), and the map $(G, \mu) \to (H, u)$ is inverse to $(H, u) \mapsto (H_{\mathbb{C}}, u_{\mathbb{C}})$ on isomorphism classes. □

Let G be a simple algebraic group $\mathbb{C}$. Choose a maximal torus T in G and a base $(\alpha_i)_{i\in I}$ for the roots of G relative to T. Recall, that the nodes of the dynkin diagram of (G, T) are indexed by I. Recall also (Bourbaki 1981, VI 1.8) that there is a unique ***(highest) root*** $\tilde{\alpha} = \sum n_i \alpha_i$ such that, for any other root $\sum m_i \alpha_i$, $n_i \geq m_i$ all i. An α_i (or the associated node) is said to be ***special*** if $n_i = 1$.

Let M be a conjugacy class of nontrivial cocharacters of G satisfying (*). Because all maximal tori of G are conjugate, M has a representative in $X_*(T) \subset X_*(G)$, and because the Weyl group acts simply transitively on the Weyl chambers (Humphreys 1972, 10.3) there is a unique representative μ for M such that

$\langle\alpha_i,\mu\rangle \geq 0$ for all $i \in I$. The condition (*) is that[4] $\langle\alpha,\mu\rangle \in \{1,0,-1\}$ for all roots α. Since μ is nontrivial, not all the values $\langle\alpha,\mu\rangle$ can be zero, and so this condition implies that $\langle\alpha_i,\mu\rangle = 1$ for exactly one $i \in I$, which must in fact be special (otherwise $\langle\tilde{\alpha},\mu\rangle > 1$). Thus, the M satisfying (*) are in one-to-one correspondence with the special nodes of the dynkin diagram. In conclusion:

THEOREM 1.25. *The isomorphism classes of irreducible hermitian symmetric domains are classified by the special nodes on connected dynkin diagrams.*

The special nodes can be read off from the list of dynkin diagrams in, for example, Helgason 1978, p477. In the following table, we list the number of special nodes for each type:

Type	A_n	B_n	C_n	D_n	E_6	E_7	E_8	F_4	G_2
	n	1	1	3	2	1	0	0	0

In particular, there are no irreducible hermitian symmetric domains of type E_8, F_4, or G_2 and, up to isomorphism, there are exactly 2 of type E_6 and 1 of type E_7. It should be noted that not every simple real algebraic group arises as the automorphism group of a hermitian symmetric domain. For example, PGL_n arises in this way only for $n = 2$.

NOTES. For introductions to smooth manifolds and riemannian manifolds, see Boothby 1975 and Lee 1997. The ultimate source for hermitian symmetric domains is Helgason 1978, but Wolf 1984 is also very useful, and Borel 1998 gives a succinct treatment close to that of the pioneers. The present account has been influenced by Deligne 1973*a* and Deligne 1979.

2. Hodge structures and their classifying spaces

We describe various objects and their parameter spaces. Our goal is a description of hermitian symmetric domains as the parameter spaces for certain special hodge structures.

Reductive groups and tensors. Let G be a reductive group over a field k of characteristic zero, and let $\rho\colon G \to \mathrm{GL}(V)$ be a representation of G. The ***contragredient*** or ***dual*** $\rho^\vee$ of ρ is the representation of G on the dual vector space $V^\vee$ defined by

$$(\rho^\vee(g)\cdot f)(v) = f(\rho(g^{-1})\cdot v), \quad g \in G,\ f \in V^\vee,\ v \in V.$$

A representation is said to be ***self-dual*** if it is isomorphic to its contragredient.

An r-***tensor*** of V is a multilinear map

$$t\colon V \times \cdots \times V \to k \quad (r\text{-copies of } V).$$

For an r-tensor t, the condition

$$t(gv_1,\ldots,gv_r) = (v_1,\ldots,v_r), \quad \text{all } v_i \in V,$$

on g defines a closed subgroup of $\mathrm{GL}(V)_t$ of $\mathrm{GL}(V)$. For example, if t is a nondegenerate symmetric bilinear form $V \times V \to k$, then $\mathrm{GL}(V)_t$ is the orthogonal group. For a set T of tensors of V, $\bigcap_{t\in T}\mathrm{GL}(V)_t$ is called the ***subgroup of*** $\mathrm{GL}(V)$ ***fixing the*** $t \in T$.

[4]The μ with this property are sometimes said to be ***minuscule*** (cf. Bourbaki 1981, pp226–227).

Proposition 2.1. *For any faithful self-dual representation $G \to \mathrm{GL}(V)$ of G, there exists a finite set T of tensors of V such that G is the subgroup of $\mathrm{GL}(V)$ fixing the $t \in T$.*

Proof. In Deligne 1982, 3.1, it is shown there exists a possibly infinite set T with this property, but, because G is noetherian as a topological space (i.e., it has the descending chain condition on closed subsets), a finite subset will suffice. □

Proposition 2.2. *Let G be the subgroup of $\mathrm{GL}(V)$ fixing the tensors $t \in T$. Then*

$$\mathrm{Lie}(G) = \left\{ g \in \mathrm{End}(V) \;\middle|\; \textstyle\sum_j t(v_1, \ldots, gv_j, \ldots, v_r) = 0, \quad \text{all } t \in T,\ v_i \in V \right\}.$$

Proof. The Lie algebra of an algebraic group G can be defined to be the kernel of $G(k[\varepsilon]) \to G(k)$. Here $k[\varepsilon]$ is the k-algebra with $\varepsilon^2 = 0$. Thus $\mathrm{Lie}(G)$ consists of the endomorphisms $1 + g\varepsilon$ of $V(k[\varepsilon])$ such that

$$t((1+g\varepsilon)v_1, (1+g\varepsilon)v_2, \ldots) = t(v_1, v_2, \ldots), \quad \text{all } t \in T,\ v_i \in V.$$

On expanding this and cancelling, we obtain the assertion. □

Flag varieties. Fix a vector space V of dimension n over a field k.

The projective space $\mathbb{P}(V)$. The set $\mathbb{P}(V)$ of one-dimensional subspaces L of V has a natural structure of an algebraic variety: the choice of a basis for V determines a bijection $\mathbb{P}(V) \to \mathbb{P}^{n-1}$, and the structure of an algebraic variety inherited by $\mathbb{P}(V)$ from the bijection is independent of the choice of the basis.

Grassmann varieties. Let $G_d(V)$ be the set of d-dimensional subspaces of V, some $0 < d < n$. Fix a basis for V. The choice of a basis for W then determines a $d \times n$ matrix $A(W)$ whose rows are the coordinates of the basis elements. Changing the basis for W multiplies $A(W)$ on the left by an invertible $d \times d$ matrix. Thus, the family of minors of degree d of $A(W)$ is well-determined up to multiplication by a nonzero constant, and so determines a point $P(W)$ in $\mathbb{P}^{\binom{n}{d}-1}$. The map $W \mapsto P(W)\colon G_d(V) \to \mathbb{P}^{\binom{n}{d}-1}$ identifies $G_d(V)$ with a closed subvariety of $\mathbb{P}^{\binom{n}{d}-1}$ (AG 6.26). A coordinate-free description of this map is given by

$$W \mapsto \textstyle\bigwedge^d W \colon G_d(V) \to \mathbb{P}(\bigwedge^d V). \tag{13}$$

Let S be a subspace of V of complementary dimension $n - d$, and let $G_d(V)_S$ be the set of $W \in G_d(V)$ such that $W \cap S = \{0\}$. Fix a $W_0 \in G_d(V)_S$, so that $V = W_0 \oplus S$. For any $W \in G_d(V)_S$, the projection $W \to W_0$ given by this decomposition is an isomorphism, and so W is the graph of a homomorphism $W_0 \to S$:

$$w \mapsto s \iff (w, s) \in W.$$

Conversely, the graph of any homomorphism $W_0 \to S$ lies in $G_d(V)_S$. Thus,

$$G_d(V)_S \cong \mathrm{Hom}(W_0, S). \tag{14}$$

When we regard $G_d(V)_S$ as an open subvariety of $G_d(V)$, this isomorphism identifies it with the affine space $\mathbb{A}(\mathrm{Hom}(W_0, S))$ defined by the vector space $\mathrm{Hom}(W_0, S)$. Thus, $G_d(V)$ is smooth, and the tangent space to $G_d(V)$ at W_0,

$$T_{W_0}(G_d(V)) \cong \mathrm{Hom}(W_0, S) \cong \mathrm{Hom}(W_0, V/W_0). \tag{15}$$

Flag varieties. The above discussion extends easily to chains of subspaces. Let $\mathbf{d} = (d_1, \dots, d_r)$ be a sequence of integers with $n > d_1 > \cdots > d_r > 0$, and let $G_{\mathbf{d}}(V)$ be the set of flags

$$F: \quad V \supset V^1 \supset \cdots \supset V^r \supset 0 \tag{16}$$

with V^i a subspace of V of dimension d_i. The map

$$G_{\mathbf{d}}(V) \xrightarrow{F \mapsto (V^i)} \prod_i G_{d_i}(V) \subset \prod_i \mathbb{P}(\textstyle\bigwedge^{d_i} V)$$

realizes $G_{\mathbf{d}}(V)$ as a closed subset of $\prod_i G_{d_i}(V)$ (Humphreys 1978, 1.8), and so it is a projective variety. The tangent space to $G_{\mathbf{d}}(V)$ at the flag F consists of the families of homomorphisms

$$\varphi^i\colon V^i \to V/V^i, \quad 1 \leq i \leq r, \tag{17}$$

satisfying the compatibility condition

$$\varphi^i | V^{i+1} \equiv \varphi^{i+1} \mod V^{i+1}.$$

Aside 2.3. A basis $e_1, \dots, e_n$ for V is ***adapted to*** the flag F if it contains a basis $e_1, \dots, e_{j_i}$ for each V^i. Clearly, every flag admits such a basis, and the basis then determines the flag. Because $\mathrm{GL}(V)$ acts transitively on the set of bases for V, it acts transitively on $G_{\mathbf{d}}(V)$. For a flag F, the subgroup $P(F)$ stabilizing F is an algebraic subgroup of $\mathrm{GL}(V)$, and the map

$$g \mapsto gF_0\colon \mathrm{GL}(V)/P(F_0) \to G_{\mathbf{d}}(V)$$

is an isomorphism of algebraic varieties. Because $G_{\mathbf{d}}(V)$ is projective, this shows that $P(F_0)$ is a parabolic subgroup of $\mathrm{GL}(V)$.

Hodge structures.

Definition. For a real vector space V, complex conjugation on $V(\mathbb{C}) =_{\mathrm{df}} \mathbb{C} \otimes_{\mathbb{R}} V$ is defined by

$$\overline{z \otimes v} = \overline{z} \otimes v.$$

An $\mathbb{R}$-basis $e_1, \dots, e_m$ for V is also a $\mathbb{C}$-basis for $V(\mathbb{C})$ and $\overline{\sum a_i e_i} = \sum \overline{a_i} e_i$.

A ***hodge decomposition*** of a real vector space V is a decomposition

$$V(\mathbb{C}) = \bigoplus_{p,q \in \mathbb{Z} \times \mathbb{Z}} V^{p,q}$$

such that $V^{q,p}$ is the complex conjugate of $V^{p,q}$. A ***hodge structure*** is a real vector space together with a hodge decomposition. The set of pairs (p, q) for which $V^{p,q} \neq 0$ is called the ***type*** of the hodge structure. For each n, $\bigoplus_{p+q=n} V^{p,q}$ is stable under complex conjugation, and so is defined over $\mathbb{R}$, i.e., there is a subspace V_n of V such that $V_n(\mathbb{C}) = \bigoplus_{p+q=n} V^{p,q}$ (see AG 16.7). Then $V = \bigoplus_n V_n$ is called the ***weight decomposition*** of V. If $V = V_n$, then V is said to have weight n.

An ***integral*** (resp. ***rational***) ***hodge structure*** is a free $\mathbb{Z}$-module of finite rank V (resp. $\mathbb{Q}$-vector space) together with a hodge decomposition of $V(\mathbb{R})$ such that the weight decomposition is defined over $\mathbb{Q}$.

Example 2.4. Let J be a complex structure on a real vector space V, and define $V^{-1,0}$ and $V^{0,-1}$ to be the $+i$ and $-i$ eigenspaces of J acting on $V(\mathbb{C})$. Then $V(\mathbb{C}) = V^{-1,0} \oplus V^{0,-1}$ is a hodge structure of type $(-1, 0), (0, -1)$, and every real hodge structure of this type arises from a (unique) complex structure. Thus, to give a rational hodge structure of type $(-1, 0), (0, -1)$ amounts to giving a $\mathbb{Q}$-vector space V and a complex structure on $V(\mathbb{R})$, and to give an integral hodge structure of type $(-1,0),(0,-1)$ amounts to giving a $\mathbb{C}$-vector space V and a lattice $\Lambda\subset V$ (i.e., a $\mathbb{Z}$-submodule generated by an $\mathbb{R}$-basis for V).

EXAMPLE 2.5. Let X be a nonsingular projective algebraic variety over $\mathbb{C}$. Then $H=H^n(X,\mathbb{Q})$ has a hodge structure of weight n for which $H^{p,q}\subset H^n(X,\mathbb{C})$ is canonically isomorphic to $H^q(X,\Omega^p)$ (Voisin 2002, 6.1.3).

EXAMPLE 2.6. Let $\mathbb{Q}(m)$ be the hodge structure of weight $-2m$ on the vector space $\mathbb{Q}$. Thus, $(\mathbb{Q}(m))(\mathbb{C})=\mathbb{Q}(m)^{-m,-m}$. Define $\mathbb{Z}(m)$ and $\mathbb{R}(m)$ similarly.[5]

The hodge filtration. The ***hodge filtration*** associated with a hodge structure of weight n is

$$F^\bullet:\quad \cdots\supset F^p\supset F^{p+1}\supset\cdots,\quad F^p=\bigoplus\nolimits_{r\geq p}V^{r,s}\subset V(\mathbb{C}).$$

Note that for $p+q=n$,

$$\overline{F^q}=\bigoplus\nolimits_{s\geq q}\overline{V^{s,r}}=\bigoplus\nolimits_{s\geq q}V^{r,s}=\bigoplus\nolimits_{r\leq p}V^{r,s}$$

and so

$$V^{p,q}=F^p\cap\overline{F^q}. \tag{18}$$

EXAMPLE 2.7. For a hodge structure of type $(-1,0),(0,-1)$, the hodge filtration is

$$(F^{-1}\supset F^0\supset F^2)=(V(\mathbb{C})\supset V^{0,-1}\supset 0).$$

The obvious $\mathbb{R}$-linear isomorphism $V\to V(\mathbb{C})/F^0$ defines the complex structure on V noted in (2.4).

Hodge structures as representations of $\mathbb{S}$. Let $\mathbb{S}$ be $\mathbb{C}^\times$ regarded as a torus over $\mathbb{R}$. It can be identified with the closed subgroup of $\mathrm{GL}_2(\mathbb{R})$ of matrices of the form $\left(\begin{smallmatrix}a&b\\-b&a\end{smallmatrix}\right)$. Then $\mathbb{S}(\mathbb{C})\approx\mathbb{C}^\times\times\mathbb{C}^\times$ with complex conjugation acting by the rule $\overline{(z_1,z_2)}=(\overline{z_2},\overline{z_1})$. We fix the isomorphism $\mathbb{S}_\mathbb{C}\cong\mathbb{G}_m\times\mathbb{G}_m$ so that $\mathbb{S}(\mathbb{R})\to\mathbb{S}(\mathbb{C})$ is $z\mapsto(z,\overline{z})$, and we define the weight homomorphism $w\colon\mathbb{G}_m\to\mathbb{S}$ so that $\mathbb{G}_m(\mathbb{R})\xrightarrow{w}\mathbb{S}(\mathbb{R})$ is $r\mapsto r^{-1}\colon\mathbb{R}^\times\to\mathbb{C}^\times$.

The characters of $\mathbb{S}_\mathbb{C}$ are the homomorphisms $(z_1,z_2)\mapsto z_1^pz_2^q$, $(r,s)\in\mathbb{Z}\times\mathbb{Z}$. Thus, $X^*(\mathbb{S})=\mathbb{Z}\times\mathbb{Z}$ with complex conjugation acting as $(p,q)\mapsto(q,p)$, and to give a representation of $\mathbb{S}$ on a real vector space V amounts to giving a $\mathbb{Z}\times\mathbb{Z}$-grading of $V(\mathbb{C})$ such that $\overline{V^{p,q}}=V^{q,p}$ for all p,q (see p276). Thus, to give a representation of $\mathbb{S}$ on a real vector space V is the same as to give a hodge structure on V. Following Deligne 1979, 1.1.1.1, we normalize the relation as follows: the homomorphism $h\colon\mathbb{S}\to\mathrm{GL}(V)$ corresponds to the hodge structure on V such that

$$h_\mathbb{C}(z_1,z_2)v=z_1^{-p}z_2^{-q}v\text{ for }v\in V^{p,q}. \tag{19}$$

In other words,

$$h(z)v=z^{-p}\overline{z}^{-q}v\text{ for }v\in V^{p,q}. \tag{20}$$

Note the minus signs! The associated weight decomposition has

$$V_n=\{v\in V\mid w_h(r)v=r^n\},\quad w_h=h\circ w. \tag{21}$$

[5]It would be a little more canonical to take the underlying vector space of $\mathbb{Q}(m)$ to be $(2\pi i)^m\mathbb{Q}$ because this makes certain relations invariant under a change of the choice of $i=\sqrt{-1}$ in $\mathbb{C}$.

Let μ_h be the cocharacter of $\mathrm{GL}(V)$ defined by

$$\mu_h(z) = h_{\mathbb{C}}(z, 1). \tag{22}$$

Then the elements of $F_h^p V$ are sums of $v \in V(\mathbb{C})$ satisfying $\mu_h(z)\, v = z^{-r} v$ for some $r \geq p$.

To give a hodge structure on a $\mathbb{Q}$-vector space V amounts to giving a homomorphism $h\colon \mathbb{S} \to \mathrm{GL}(V(\mathbb{R}))$ such that w_h is defined over $\mathbb{Q}$.

EXAMPLE 2.8. By definition, a complex structure on a real vector space is a homomorphism $h\colon \mathbb{C} \to \mathrm{End}_{\mathbb{R}}(V)$ of $\mathbb{R}$-algebras. Then $h|\mathbb{C}^\times\colon \mathbb{C}^\times \to \mathrm{GL}(V)$ is a hodge structure of type $(-1,0),(0,-1)$ whose associated complex structure (see 2.4) is that defined by h.[6]

EXAMPLE 2.9. The hodge structure $\mathbb{Q}(m)$ corresponds to the homomorphism $h\colon \mathbb{S} \to \mathbb{G}_{m\mathbb{R}}$, $h(z) = (z\overline{z})^m$.

The Weil operator. For a hodge structure (V, h), the $\mathbb{R}$-linear map $C = h(i)$ is called the ***Weil operator***. Note that C acts as i^{q-p} on $V^{p,q}$ and that $C^2 = h(-1)$ acts as $(-1)^n$ on V_n.

EXAMPLE 2.10. If V is of type $(-1,0)$, $(0,-1)$, then C coincides with the J of (2.4). The functor $(V, (V^{-1,0}, V^{0,-1})) \mapsto (V, C)$ is an equivalence from the category of real hodge structures of type $(-1,0),(0,-1)$ to the category of complex vector spaces.

Hodge structures of weight 0.. Let V be a hodge structure of weight 0. Then $V^{0,0}$ is invariant under complex conjugation, and so $V^{0,0} = V^{00}(\mathbb{C})$, where $V^{00} = V^{0,0} \cap V$ (see AG 16.7). Note that

$$V^{00} = \mathrm{Ker}(V \to V(\mathbb{C})/F^0). \tag{23}$$

Tensor products of hodge structures. The ***tensor product of hodge structures*** V and W of weight m and n is a hodge structure of weight $m+n$:

$$V \otimes W, \quad (V \otimes W)^{p,q} = \bigoplus\nolimits_{r+r'=p,s+s'=q} V^{r,s} \otimes V^{r',s'}.$$

In terms of representations of $\mathbb{S}$,

$$(V, h_V) \otimes (W, h_W) = (V \otimes W, h_V \otimes h_W).$$

Morphisms of hodge structures. A morphism of hodge structures is a linear map $V \to W$ sending $V^{p,q}$ into $W^{p,q}$ for all p, q. In other words, it is a morphism $(V, h_V) \to (W, h_W)$ of representations of $\mathbb{S}$.

Hodge tensors. Let $R = \mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$, and let (V, h) be an R-hodge structure of weight n. A multilinear form $t\colon V^r \to R$ is a ***hodge tensor*** if the map

$$V \otimes V \otimes \cdots \otimes V \to R(-nr/2)$$

it defines is a morphism of hodge structures. In other words, t is a hodge tensor if

$$t(h(z)v_1, h(z)v_2, \ldots) = (z\overline{z})^{-nr/2} \cdot t_{\mathbb{R}}(v_1, v_2, \ldots), \text{ all } z \in \mathbb{C},\ v_i \in V(\mathbb{R}),$$

or if

$$\textstyle\sum p_i \neq \sum q_i \Rightarrow t_{\mathbb{C}}(v_1^{p_1,q_1}, v_2^{p_2,q_2}, \ldots) = 0, \quad v_i^{p_i,q_i} \in V^{p_i,q_i}. \tag{24}$$

[6] This partly explains the signs in (19); see also Deligne 1979, 1.1.6. Following Deligne 1973*b*, 8.12, and Deligne 1979, 1.1.1.1, $h_{\mathbb{C}}(z_1, z_2)v^{p,q} = z_1^{-p} z_2^{-q} v^{p,q}$ has become the standard convention in the theory of Shimura varieties. Following Deligne 1971*a*, 2.1.5.1, the convention $h_{\mathbb{C}}(z_1, z_2)v^{p,q} = z_1^p z_2^q v^{p,q}$ is commonly used in hodge theory (e.g., Voisin 2002, p147).

Note that, for a hodge tensor t,

$$t(Cv_1, Cv_2, \ldots) = t(v_1, v_2, \ldots).$$

EXAMPLE 2.11. Let (V, h) be a hodge structure of type $(-1, 0), (0, -1)$. A bilinear form $t\colon V \times V \to \mathbb{R}$ is a hodge tensor if and only if $t(Ju, Jv) = t(u, v)$ for all $u, v \in V$.

Polarizations. Let (V, h) be a hodge structure of weight n. A ***polarization*** of (V, h) is a hodge tensor $\psi\colon V \times V \to \mathbb{R}$ such that $\psi_C(u, v) =_{\mathrm{df}} \psi(u, Cv)$ is symmetric and positive definite. Then ψ is symmetric or alternating according as n is even or odd, because

$$\psi(v, u) = \psi(Cv, Cu) = \psi_C(Cv, u) = \psi_C(u, Cv) = \psi(u, C^2 v) = (-1)^n \psi(u, v).$$

More generally, let (V, h) be an R-hodge structure of weight n where R is $\mathbb{Z}$ or $\mathbb{Q}$. A ***polarization*** of (V, h) is a bilinear form $\psi\colon V \times V \to R$ such that $\psi_{\mathbb{R}}$ is a polarization of $(V(\mathbb{R}), h)$.

EXAMPLE 2.12. Let (V, h) be an R-hodge structure of type $(-1, 0), (0, -1)$ with $R = \mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$, and let $J = h(i)$. A polarization of (V, h) is an alternating bilinear form $\psi\colon V \times V \to R$ such that, for $u, v \in V(\mathbb{R})$,

$$\psi_{\mathbb{R}}(Ju, Jv) = \psi(u, v), \text{ and}$$
$$\psi_{\mathbb{R}}(u, Ju) > 0 \text{ if } u \neq 0.$$

(These conditions imply that $\psi_{\mathbb{R}}(u, Jv)$ is symmetric.)

EXAMPLE 2.13. Let X be a nonsingular projective variety over $\mathbb{C}$. The choice of an embedding $X \hookrightarrow \mathbb{P}^N$ determines a polarization on the primitive part of $H^n(X, \mathbb{Q})$ (Voisin 2002, 6.3.2).

Variations of hodge structures. Fix a real vector space V, and let S be a connected complex manifold. Suppose that, for each $s \in S$, we have a hodge structure h_s on V of weight n (independent of s). Let $V_s^{p,q} = V_{h_s}^{p,q}$ and $F_s^p = F_s^p V = F_{h_s}^p V$.

The family of hodge structures $(h_s)_{s \in S}$ on V is said to be ***continuous*** if, for fixed p and q, the subspace $V_s^{p,q}$ varies continuously with s. This means that the dimension $d(p, q)$ of $V_s^{p,q}$ is constant and the map

$$s \mapsto V_s^{p,q}\colon S \to G_{d(p,q)}(V)$$

is continuous.

A continuous family of hodge structures $(V_s^{p,q})_s$ is said to be ***holomorphic*** if the hodge filtration $F_s^\bullet$ varies holomorphically with s. This means that the map φ,

$$s \mapsto F_s^\bullet\colon S \to G_{\mathbf{d}}(V)$$

is holomorphic. Here $\mathbf{d} = (\ldots, d(p), \ldots)$ where $d(p) = \dim F_s^p V = \sum_{r \geq p} d(r, q)$. Then the differential of φ at s is a $\mathbb{C}$-linear map

$$d\varphi_s\colon T_s S \to T_{F_s^\bullet}(G_{\mathbf{d}}(V)) \overset{(17)}{\subset} \bigoplus\nolimits_p \operatorname{Hom}(F_s^p, V/F_s^p).$$

If the image of $d\varphi_s$ is contained in

$$\bigoplus\nolimits_p \operatorname{Hom}(F_s^p, F_s^{p-1}/F_s^p),$$

for all s, then the holomorphic family is called a ***variation of hodge structures on*** S.

Now let T be a family of tensors on V including a nondegenerate bilinear form t_0, and let $d\colon \mathbb{Z}\times\mathbb{Z}\to\mathbb{N}$ be a function such that

$$\begin{aligned} &d(p,q)=0 \text{ for almost all } p,q;\\ &d(q,p)=d(p,q);\\ &d(p,q)=0 \text{ unless } p+q=n. \end{aligned}$$

Define $S(d,T)$ to be the set of all hodge structures h on V such that

- $\dim V_h^{p,q}=d(p,q)$ for all p,q;
- each $t\in T$ is a hodge tensor for h;
- t_0 is a polarization for h.

Then $S(d,T)$ acquires a topology as a subspace of $\prod_{d(p,q)\neq 0} G_{d(p,q)}(V)$.

THEOREM 2.14. *Let S^+ be a connected component of $S(d,T)$.*

(a) *If nonempty, S^+ has a unique complex structure for which (h_s) is a holomorphic family of hodge structures.*
(b) *With this complex structure, S^+ is a hermitian symmetric domain if (h_s) is a variation of hodge structures.*
(c) *Every irreducible hermitian symmetric domain is of the form S^+ for a suitable V, d, and T.*

PROOF (SKETCH). (a) Let $S^+=S(d,T)^+$. Because the hodge filtration determines the hodge decomposition (see (18)), the map $x\mapsto F_s^\bullet\colon S^+\xrightarrow{\varphi} G_{\mathbf{d}}(V)$ is injective. Let G be the smallest algebraic subgroup of $\mathrm{GL}(V)$ such that

$$h(\mathbb{S})\subset G,\quad \text{all } h\in S^+ \tag{25}$$

(take G to be the intersection of the algebraic subgroups of $\mathrm{GL}(V)$ with this property), and let $h_o\in S^+$. For any $g\in G(\mathbb{R})^+$, $gh_og^{-1}\in S^+$, and it can be shown that the map $g\mapsto g\cdot h_o\cdot g^{-1}\colon G(\mathbb{R})^+\to S^+$ is surjective:

$$S^+=G(\mathbb{R})^+\cdot h_o.$$

The subgroup K_o of $G(\mathbb{R})^+$ fixing h_o is closed, and so $G(\mathbb{R})^+/K_o$ is a smooth (in fact, real analytic) manifold. Therefore, S^+ acquires the structure of a smooth manifold from

$$S^+=(G(\mathbb{R})^+/K_o)\cdot h_o\cong G(\mathbb{R})^+/K_o.$$

Let $\mathfrak{g}=\mathrm{Lie}(G)$. From $\mathbb{S}\xrightarrow{h_o}G\xrightarrow{\mathrm{Ad}}\mathfrak{g}\subset\mathrm{End}(V)$, we obtain hodge structures on $\mathfrak{g}$ and $\mathrm{End}(V)$. Clearly, $\mathfrak{g}^{00}=\mathrm{Lie}(K_o)$ and so $T_{h_o}S^+\cong\mathfrak{g}/\mathfrak{g}^{00}$. In the diagram,

$$\begin{tikzcd} T_{h_o}S^+\cong \mathfrak{g}/\mathfrak{g}^{00} \arrow[r, hook] \arrow[d, "(23)"', "\cong"] & \mathrm{End}(V)/\mathrm{End}(V)^{00} \arrow[d, "(23)"', "\cong"] \\ \mathfrak{g}(\mathbb{C})/F^0 \arrow[r, hook] & \mathrm{End}(V(\mathbb{C}))/F^0 \cong T_{h_o}G_{\mathbf{d}}(V). \end{tikzcd} \tag{26}$$

the map from top-left to bottom-right is $(d\varphi)_{h_o}$, which therefore maps $T_{h_o}S^+$ onto a complex subspace of $T_{h_o}G_{\mathbf{d}}(V)$. Since this is true for all $h_o\in S^+$, we see that φ identifies S^+ with an almost-complex submanifold $G_{\mathbf{d}}(V)$. It can be shown that this almost-complex structure is integrable, and so provides S^+ with a complex structure for which φ is holomorphic. Clearly, this is the only (almost-)complex structure for which this is true.

(b) See Deligne 1979, 1.1.

(c) Given an irreducible hermitian symmetric domain D, choose a faithful self-dual representation $G \to \mathrm{GL}(V)$ of the algebraic group G associated with D (as in 1.7). Because V is self-dual, there is a nondegenerate bilinear form t_0 on V fixed by G. Apply Theorem 2.1 to find a set of tensors T such that G is the subgroup of $\mathrm{GL}(V)$ fixing the $t \in T$. Let h_o be the composite $\mathbb{S} \xrightarrow{z \mapsto z/\bar{z}} U_1 \xrightarrow{u_o} \mathrm{GL}(V)$ with u_o as in (1.9). Then, h_o defines a hodge structure on V for which the $t \in T$ are hodge tensors and t_o is a polarization. One can check that D is naturally identified with the component of $S(d,T)^+$ containing this hodge structure. □

Remark 2.15. The map $S^+ \to G_{\mathbf{d}}(V)$ in the proof is an embedding of smooth manifolds (injective smooth map that is injective on tangent spaces and maps S^+ homeomorphically onto its image). Therefore, if a smooth map $T \to G_{\mathbf{d}}(V)$ factors into

$$T \xrightarrow{\alpha} S^+ \longrightarrow G_{\mathbf{d}}(V),$$

then α will be smooth. Moreover, if the map $T \to G_{\mathbf{d}}(V)$ is defined by a holomorphic family of hodge structures on T, and it factors through S^+, then α will be holomorphic.

Aside 2.16. As we noted in (2.5), for a nonsingular projective variety V over $\mathbb{C}$, the cohomology group $H^n(V(\mathbb{C}),\mathbb{Q})$ has a natural hodge structure of weight n. Now consider a regular map $\pi\colon V \to S$ of nonsingular varieties whose fibres V_s ($s \in S$) are nonsingular projective varieties of constant dimension. The vector spaces $H^n(V_s,\mathbb{Q})$ form a local system of $\mathbb{Q}$-vector spaces on S, and Griffiths showed that the hodge structures on them form a variation of hodge structures in a slightly more general sense than that defined above (Voisin 2002, Proposition 10.12).

Notes. Theorem 2.14 is taken from Deligne 1979.

3. Locally symmetric varieties

In this section, we study quotients of hermitian symmetric domains by certain discrete groups.

Quotients of hermitian symmetric domains by discrete groups.

Proposition 3.1. *Let D be a hermitian symmetric domain, and let Γ be a discrete subgroup of* $\mathrm{Hol}(D)^+$. *If Γ is torsion free, then Γ acts freely on D, and there is a unique complex structure on $\Gamma\backslash D$ for which the quotient map $\pi\colon D \to \Gamma\backslash D$ is a local isomorphism. Relative to this structure, a map φ from $\Gamma\backslash D$ to a second complex manifold is holomorphic if and only if $\varphi\circ\pi$ is holomorphic.*

Proof. Let Γ be a discrete subgroup of $\mathrm{Hol}(D)^+$. According to (1.5, 1.6), the stabilizer K_p of any point $p \in D$ is compact and $g \mapsto gp\colon \mathrm{Hol}(D)^+/K_p \to D$ is a homeomorphism, and so (MF, 2.5):

(a) for any $p \in D$, $\{g \in \Gamma \mid gp = p\}$ is finite;
(b) for any $p \in D$, there exists a neighbourhood U of p such that, for $g \in \Gamma$, gU is disjoint from U unless $gp = p$;
(c) for any points $p, q \in D$ not in the same Γ-orbit, there exist neighbourhoods U of p and V of q such that $gU \cap V = \emptyset$ for all $g \in \Gamma$.

Assume Γ is torsion free. Then the group in (a) is trivial, and so Γ acts freely on D. Endow $\Gamma\backslash D$ with the quotient topology. If U and V are as in (c) , then πU and πV are disjoint neighbourhoods of πp and πq, and so $\Gamma\backslash D$ is separated. Let $q \in \Gamma\backslash D$, and let $p \in \pi^{-1}(q)$. If U is as in (b), then the restriction of π to U is a homeomorphism $U \to \pi U$, and it follows that $\Gamma\backslash D$ a manifold.

Define a $\mathbb{C}$-valued function f on an open subset U of $\Gamma\backslash D$ to be holomorphic if $f \circ \pi$ is holomorphic on $\pi^{-1}U$. The holomorphic functions form a sheaf on $\Gamma\backslash D$ for which π is a local isomorphism of ringed spaces. Therefore, the sheaf defines a complex structure on $\Gamma\backslash D$ for which π is a local isomorphism of complex manifolds.

Finally, let $\varphi\colon \Gamma\backslash D \to M$ be a map such that $\varphi\circ\pi$ is holomorphic, and let f be a holomorphic function on an open subset U of M. Then $f\circ\varphi$ is holomorphic because $f\circ\varphi\circ\pi$ is holomorphic, and so φ is holomorphic. $\square$

When Γ is torsion free, we often write $D(\Gamma)$ for $\Gamma\backslash D$ regarded as a complex manifold. In this case, D is the universal covering space of $D(\Gamma)$ and Γ is the group of covering transformations; moreover, for any point p of D, the map

$$g \mapsto [\text{image under } \pi \text{ of any path from } p \text{ to } gp]\colon \Gamma \to \pi_1(D(\Gamma), \pi p)$$

is an isomorphism (Hatcher 2002, 1.40).

Subgroups of finite covolume. We shall only be interested in quotients of D by "big" discrete subgroups Γ of $\mathrm{Aut}(D)^+$. This condition is conveniently expressed by saying that $\Gamma\backslash D$ has finite volume. By definition, D has a riemannian metric g and hence a volume element Ω: in local coordinates

$$\Omega = \sqrt{\det(g_{ij}(x))}dx^1 \wedge \ldots \wedge dx^n.$$

Since g is invariant under Γ, so also is Ω, and so it passes to the quotient $\Gamma\backslash D$. The condition is that $\int_{\Gamma\backslash D}\Omega < \infty$.

For example, let $D = \mathcal{H}_1$ and let $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. Then

$$F = \{z \in \mathcal{H}_1 \mid |z| > 1, \quad -\tfrac{1}{2} < \Re z < \tfrac{1}{2}\}$$

is a fundamental domain for Γ and

$$\int_{\Gamma\backslash D}\Omega = \iint_F \frac{dxdy}{y^2} \le \int_{\sqrt{3}/2}^{\infty}\int_{-1/2}^{1/2}\frac{dxdy}{y^2} = \int_{\sqrt{3}/2}^{\infty}\frac{dy}{y^2} < \infty.$$

On the other hand, the quotient of $\mathcal{H}_1$ by the group of translations $z \mapsto z+n$, $n \in \mathbb{Z}$, has infinite volume, as does the quotient of $\mathcal{H}_1$ by the trivial group.

A real Lie group G has a left invariant volume element, which is unique up to a positive constant (cf. Boothby 1975, VI 3.5). A discrete subgroup Γ of G is said to have ***finite covolume*** if $\Gamma\backslash G$ has finite volume. For a torsion free discrete subgroup Γ of $\mathrm{Hol}(D)^+$, an application of Fubini's theorem shows that $\Gamma\backslash\mathrm{Hol}(D)^+$ has finite volume if and only if $\Gamma\backslash D$ has finite volume (Witte 2001, Exercise 1.27).

Arithmetic subgroups. Two subgroups S_1 and S_2 of a group H are ***commensurable*** if $S_1 \cap S_2$ has finite index in both S_1 and S_2. For example, two infinite cyclic subgroups $\mathbb{Z}a$ and $\mathbb{Z}b$ of $\mathbb{R}$ are commensurable if and only if $a/b \in \mathbb{Q}^\times$. Commensurability is an equivalence relation.

Let G be an algebraic group over $\mathbb{Q}$. A subgroup Γ of $G(\mathbb{Q})$ is ***arithmetic*** if it is commensurable with $G(\mathbb{Q}) \cap \mathrm{GL}_n(\mathbb{Z})$ for some embedding $G \hookrightarrow \mathrm{GL}_n$. It is then commensurable with $G(\mathbb{Q}) \cap \mathrm{GL}_{n'}(\mathbb{Z})$ for every embedding $G \hookrightarrow \mathrm{GL}_{n'}$ (Borel 1969, 7.13).

PROPOSITION 3.2. *Let $\rho\colon G \to G'$ be a surjective homomorphism of algebraic groups over $\mathbb{Q}$. If $\Gamma \subset G(\mathbb{Q})$ is arithmetic, then so also is $\rho(\Gamma) \subset G'(\mathbb{Q})$.*

PROOF. Borel 1969, 8.9, 8.11, or Platonov and Rapinchuk 1994, Theorem 4.1, p204. □

An arithmetic subgroup Γ of $G(\mathbb{Q})$ is obviously discrete in $G(\mathbb{R})$, but it need not have finite covolume; for example, $\Gamma = \{\pm 1\}$ is an arithmetic subgroup of $\mathbb{G}_m(\mathbb{Q})$ of infinite covolume in $\mathbb{R}^\times$. Thus, if Γ is to have finite covolume, there can be no nonzero homomorphism $G \to \mathbb{G}_m$. For reductive groups, this condition is also sufficient.

THEOREM 3.3. *Let G be a reductive group over $\mathbb{Q}$, and let Γ be an arithmetic subgroup of $G(\mathbb{Q})$.*

(a) *The space $\Gamma\backslash G(\mathbb{R})$ has finite volume if and only if* $\mathrm{Hom}(G, \mathbb{G}_m) = 0$ *(in particular, $\Gamma\backslash G(\mathbb{R})$ has finite volume if G is semisimple).*[7]
(b) *The space $\Gamma\backslash G(\mathbb{R})$ is compact if and only if* $\mathrm{Hom}(G, \mathbb{G}_m) = 0$ *and $G(\mathbb{Q})$ contains no unipotent element (other than 1).*

PROOF. Borel 1969, 13.2, 8.4, or Platonov and Rapinchuk 1994, Theorem 4.13, p213, Theorem 4.12, p210. [The intuitive reason for the condition in (b) is that the rational unipotent elements correspond to cusps (at least in the case of SL_2 acting on $\mathcal{H}_1$), and so no rational unipotent elements means no cusps.] □

EXAMPLE 3.4. Let B be a quaternion algebra over $\mathbb{Q}$ such that $B \otimes_{\mathbb{Q}} \mathbb{R} \approx M_2(\mathbb{R})$, and let G be the algebraic group over $\mathbb{Q}$ such that $G(\mathbb{Q})$ is the group of elements in B of norm 1. The choice of an isomorphism $B \otimes_{\mathbb{Q}} \mathbb{R} \to M_2(\mathbb{R})$ determines an isomorphism $G(\mathbb{R}) \to \mathrm{SL}_2(\mathbb{R})$, and hence an action of $G(\mathbb{R})$ on $\mathcal{H}_1$. Let Γ be an arithmetic subgroup of $G(\mathbb{Q})$.

If $B \approx M_2(\mathbb{Q})$, then $G \approx \mathrm{SL}_2$, which is semisimple, and so $\Gamma\backslash \mathrm{SL}_2(\mathbb{R})$ (hence also $\Gamma\backslash\mathcal{H}_1$) has finite volume. However, $\mathrm{SL}_2(\mathbb{Q})$ contains the unipotent element $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$, and so $\Gamma\backslash \mathrm{SL}_2(\mathbb{R})$ is not compact.

If $B \not\approx M_2(\mathbb{Q})$, it is a division algebra, and so $G(\mathbb{Q})$ contains no unipotent element $\neq 1$ (for otherwise $B^\times$ would contain a nilpotent element). Therefore, $\Gamma\backslash G(\mathbb{R})$ (hence also $\Gamma\backslash\mathcal{H}_1$) is compact

Let k be a subfield of $\mathbb{C}$. An automorphism α of a k-vector space V is said to be ***neat*** if its eigenvalues in $\mathbb{C}$ generate a torsion free subgroup of $\mathbb{C}^\times$ (which implies that α does not have finite order). Let G be an algebraic group over $\mathbb{Q}$. An element $g \in G(\mathbb{Q})$ is ***neat*** if $\rho(g)$ is neat for one faithful representation $G \hookrightarrow \mathrm{GL}(V)$, in which case $\rho(g)$ is neat for every representation ρ of G defined over a subfield of $\mathbb{C}$ (apply Waterhouse 1979, 3.5). A subgroup of $G(\mathbb{Q})$ is ***neat*** if all its elements are.

PROPOSITION 3.5. *Let G be an algebraic group over $\mathbb{Q}$, and let Γ be an arithmetic subgroup of $G(\mathbb{Q})$. Then, Γ contains a neat subgroup Γ' of finite index. Moreover, Γ' can be defined by congruence conditions (i.e., for some embedding $G \hookrightarrow \mathrm{GL}_n$ and integer N, $\Gamma' = \{g \in \Gamma \mid g \equiv 1 \bmod N\}$).*

[7] Recall (cf. the Notations) that $\mathrm{Hom}(G, \mathbb{G}_m) = 0$ means that there is no nonzero homomorphism $G \to \mathbb{G}_m$ *defined over* $\mathbb{Q}$.

PROOF. Borel 1969, 17.4. □

Let H be a connected real Lie group. A subgroup Γ of H is ***arithmetic*** if there exists an algebraic group G over $\mathbb{Q}$ and an arithmetic subgroup Γ_0 of $G(\mathbb{Q})$ such that $\Gamma_0 \cap G(\mathbb{R})^+$ maps onto Γ under a surjective homomorphism $G(\mathbb{R})^+ \to H$ with compact kernel.

PROPOSITION 3.6. *Let H be a semisimple real Lie group that admits a faithful finite-dimensional representation. Every arithmetic subgroup Γ of H is discrete of finite covolume, and it contains a torsion free subgroup of finite index.*

PROOF. Let $\alpha\colon G(\mathbb{R})^+ \twoheadrightarrow H$ and $\Gamma_0 \subset G(\mathbb{Q})$ be as in the definition of arithmetic subgroup. Because $\mathrm{Ker}(\alpha)$ is compact, α is proper (Bourbaki 1989, I 10.3) and, in particular, closed. Because Γ_0 is discrete in $G(\mathbb{R})$, there exists an open $U \subset G(\mathbb{R})^+$ whose intersection with Γ_0 is exactly the kernel of $\Gamma_0 \cap G(\mathbb{R})^+ \to \Gamma$. Now $\alpha(G(\mathbb{R})^+ \smallsetminus U)$ is closed in H, and its complement intersects Γ in $\{1_\Gamma\}$. Therefore, Γ is discrete in H. It has finite covolume because $\Gamma_0 \backslash G(\mathbb{R})^+$ maps onto $\Gamma \backslash H$ and we can apply (3.3a). Let Γ_1 be a neat subgroup of Γ_0 of finite index (3.5). The image of Γ_1 in H has finite index in Γ, and its image under any faithful representation of H is torsion free. □

REMARK 3.7. There are many nonarithmetic discrete subgroup in $\mathrm{SL}_2(\mathbb{R})$ of finite covolume. According to the Riemann mapping theorem, every compact riemann surface of genus $g \geq 2$ is the quotient of $\mathcal{H}_1$ by a discrete subgroup of $\mathrm{PGL}_2(\mathbb{R})^+$ acting freely on $\mathcal{H}_1$. Since there are continuous families of such riemann surfaces, this shows that there are uncountably many discrete cocompact subgroups in $\mathrm{PGL}_2(\mathbb{R})^+$ (therefore also in $\mathrm{SL}_2(\mathbb{R})$), but there only countably many arithmetic subgroups.

The following (Fields medal) theorem of Margulis shows that SL_2 is exceptional in this regard: let Γ be a discrete subgroup of finite covolume in a noncompact simple real Lie group H; then Γ is arithmetic unless H is isogenous to $\mathrm{SO}(1,n)$ or $\mathrm{SU}(1,n)$ (see Witte 2001, 6.21 for a discussion of the theorem). Note that, because $\mathrm{SL}_2(\mathbb{R})$ is isogenous to $\mathrm{SO}(1,2)$, the theorem doesn't apply to it.

Brief review of algebraic varieties. Let k be a field. An ***affine k-algebra*** is a finitely generated k-algebra A such that $A \otimes_k k^{\mathrm{al}}$ is reduced (i.e., has no nilpotents). Such an algebra is itself reduced, and when k is perfect every reduced finitely generated k-algebra is affine.

Let A be an affine k-algebra. Define $\mathrm{specm}(A)$ to be the set of maximal ideals in A endowed with the topology having as basis $D(f)$, $D(f) = \{\mathfrak{m} \mid f \notin \mathfrak{m}\}$, $f \in A$. There is a unique sheaf of k-algebras $\mathcal{O}$ on $\mathrm{specm}(A)$ such that $\mathcal{O}(D(f)) = A_f$ for all f. Here A_f is the algebra obtained from A by inverting f. Any ringed space isomorphic to a ringed space of the form

$$\mathrm{Specm}(A) = (\mathrm{specm}(A), \mathcal{O})$$

is called an ***affine variety*** over k. The stalk at $\mathfrak{m}$ is the local ring $A_{\mathfrak{m}}$, and so $\mathrm{Specm}(A)$ is a locally ringed space.

This all becomes much more familiar when k is algebraically closed. When we write $A = k[X_1, \ldots, X_n]/\mathfrak{a}$, the space $\mathrm{specm}(A)$ becomes identified with the zero set of $\mathfrak{a}$ in k^n endowed with the zariski topology, and $\mathcal{O}$ becomes identified with the sheaf of k-valued functions on $\mathrm{specm}(A)$ locally defined by polynomials.

A topological space V with a sheaf of k-algebras $\mathcal{O}$ is a ***prevariety*** over k if there exists a finite covering (U_i) of V by open subsets such that $(U_i, \mathcal{O}|U_i)$ is an affine variety over k for all i. A ***morphism of prevarieties over*** k is simply a morphism of ringed spaces of k-algebras. A prevariety V over k is ***separated*** if, for all pairs of morphisms of k-prevarieties $\alpha, \beta \colon Z \rightrightarrows V$, the subset of Z on which α and β agree is closed. A ***variety over*** k is a separated prevariety over k.

Alternatively, the varieties over k are precisely the ringed spaces obtained from geometrically-reduced separated schemes of finite type over k by deleting the nonclosed points.

A morphism of algebraic varieties is also called a ***regular map***, and the elements of $\mathcal{O}(U)$ are called the ***regular functions*** on U.

For the variety approach to algebraic geometry, see AG, and for the scheme approach, see Hartshorne 1977.

Algebraic varieties versus complex manifolds.

The functor from nonsingular algebraic varieties to complex manifolds. For a nonsingular variety V over $\mathbb{C}$, $V(\mathbb{C})$ has a natural structure as a complex manifold. More precisely:

PROPOSITION 3.8. *There is a unique functor $(V, \mathcal{O}_V) \mapsto (V^{an}, \mathcal{O}_{V^{an}})$ from nonsingular varieties over $\mathbb{C}$ to complex manifolds with the following properties:*

(a) *as sets, $V = V^{an}$, every zariski-open subset is open for the complex topology, and every regular function is holomorphic;*
(b) *if $V = \mathbb{A}^n$, then $V^{an} = \mathbb{C}^n$ with its natural structure as a complex manifold;*
(c) *if $\varphi \colon V \to W$ is étale, then $\varphi^{an} \colon V^{an} \to W^{an}$ is a local isomorphism.*

PROOF. A regular map $\varphi \colon V \to W$ is étale if the map $d\varphi_p \colon T_pV \to T_pW$ is an isomorphism for all $p \in V$. Note that conditions (a,b,c) determine the complex-manifold structure on any open subvariety of $\mathbb{A}^n$ and also on any variety V that admits an étale map to an open subvariety of $\mathbb{A}^n$. Since every nonsingular variety admits a zariski-open covering by such V (AG 5.27), this shows that there exists at most one functor satisfying (a,b,c), and suggests how to define it. $\square$

Obviously, a regular map $\varphi \colon V \to W$ is determined by $\varphi^{\mathrm{an}} \colon V^{\mathrm{an}} \to W^{\mathrm{an}}$, but not every holomorphic map $V^{\mathrm{an}} \to W^{\mathrm{an}}$ is regular. For example, $z \mapsto e^z \colon \mathbb{C} \to \mathbb{C}$ is not regular. Moreover, a complex manifold need not arise from a nonsingular algebraic variety, and two nonsingular varieties V and W can be isomorphic as complex manifolds without being isomorphic as algebraic varieties (Shafarevich 1994, VIII 3.2). In other words, the functor $V \mapsto V^{\mathrm{an}}$ is faithful, but it is neither full nor essentially surjective on objects.

REMARK 3.9. The functor $V \mapsto V^{\mathrm{an}}$ can be extended to all algebraic varieties once one has the notion of a "complex manifold with singularities". This is called a ***complex space***. For holomorphic functions $f_1, \ldots, f_r$ on a connected open subset U of $\mathbb{C}^n$, let $V(f_1, \ldots, f_r)$ denote the set of common zeros of the f_i in U; one endows $V(f_1, \ldots, f_r)$ with a natural structure of ringed space, and then defines a complex space to be a ringed space $(S, \mathcal{O}_S)$ that is locally isomorphic to one of this form (Shafarevich 1994, VIII 1.5).

Necessary conditions for a complex manifold to be algebraic.

3.10. Here are two necessary conditions for a complex manifold M to arise from an algebraic variety.

(a) It must be possible to embed M as an open submanifold of a compact complex manfold M^* in such a way that the boundary $M^* \smallsetminus M$ is a finite union of manifolds of dimension $\dim M - 1$.
(b) If M is compact, then the field of meromorphic functions on M must have transcendence degree $\dim M$ over $\mathbb{C}$.

The necessity of (a) follows from Hironaka's theorem on the resolution of singularities, which shows that every nonsingular variety V can be embedded as an open subvariety of a complete nonsingular variety V^* in such a way that the boundary $V^* \smallsetminus V$ is a divisor with normal crossings (see p293), and the necessity of (b) follows from the fact that, when V is complete and nonsingular, the field of meromorphic functions on V^{an} coincides with the field of rational functions on V (Shafarevich 1994, VIII 3.1).

Here is one positive result: the functor

$$\{\text{projective nonsingular curves over } \mathbb{C}\} \to \{\text{compact riemann surfaces}\}$$

is an equivalence of categories (see MF, pp88-91, for a discussion of this theorem). Since the proper zariski-closed subsets of algebraic curves are the finite subsets, we see that for riemann surfaces the condition (3.10a) is also sufficient: a riemann surface M is algebraic if and only if it is possible to embed M in a compact riemann surface M^* in such a way that the boundary $M^* \smallsetminus M$ is finite. The maximum modulus principle (Cartan 1963, VI 4.4) shows that a holomorphic function on a connected compact riemann surface is constant. Therefore, if a connected riemann surface M is algebraic, then every bounded holomorphic function on M is constant. We conclude that $\mathcal{H}_1$ does not arise from an algebraic curve, because the function $z \mapsto \frac{z-i}{z+i}$ is bounded, holomorphic, and nonconstant.

For any lattice Λ in $\mathbb{C}$, the Weierstrass $\wp$ function and its derivative embed $\mathbb{C}/\Lambda$ into $\mathbb{P}^2(\mathbb{C})$ (as an elliptic curve). However, for a lattice Λ in $\mathbb{C}^2$, the field of meromorphic functions on $\mathbb{C}^2/\Lambda$ will usually have transcendence degree < 2, and so $\mathbb{C}^2/\Lambda$ is not an algebraic variety. For quotients of $\mathbb{C}^g$ by a lattice Λ, condition (3.10b) is sufficient for algebraicity (Mumford 1970, p35).

Projective manifolds and varieties. A complex manifold (resp. algebraic variety) is ***projective*** if it is isomorphic to a closed submanifold (resp. closed subvariety) of a projective space. The first truly satisfying theorem in the subject is the following:

THEOREM 3.11 (Chow 1949). *Every projective complex manifold has a unique structure of a nonsingular projective algebraic variety, and every holomorphic map of projective complex manifolds is regular for these structures. (Moreover, a similar statement holds for complex spaces.)*

PROOF. See Shafarevich 1994, VIII 3.1 (for the manifold case). □

In other words, the functor $V \mapsto V^{\mathrm{an}}$ is an equivalence from the category of (nonsingular) projective algebraic varieties to the category of projective complex (manifolds) spaces.

The theorem of Baily and Borel.

THEOREM 3.12 (Baily and Borel 1966). *Let $D(\Gamma) = \Gamma\backslash D$ be the quotient of a hermitian symmetric domain by a torsion free arithmetic subgroup Γ of $\mathrm{Hol}(D)^+$.*

Then $D(\Gamma)$ has a canonical realization as a zariski-open subset of a projective algebraic variety $D(\Gamma)^*$. In particular, it has a canonical structure as an algebraic variety.

Recall the proof for $D = \mathcal{H}_1$. Set $\mathcal{H}_1^* = \mathcal{H}_1 \cup \mathbb{P}^1(\mathbb{Q})$ (rational points on the real axis plus the point $i\infty$). Then Γ acts on $\mathcal{H}_1^*$, and the quotient $\Gamma\backslash\mathcal{H}_1^*$ is a compact riemann surface. One can then show that the modular forms of a sufficiently high weight embed $\Gamma\backslash\mathcal{H}_1^*$ as a closed submanifold of a projective space. Thus $\Gamma\backslash\mathcal{H}_1^*$ is algebraic, and as $\Gamma\backslash\mathcal{H}_1$ omits only finitely many points of $\Gamma\backslash\mathcal{H}_1^*$, it is automatically a zariski-open subset of $\Gamma\backslash\mathcal{H}_1^*$. The proof in the general case is similar, but is much more difficult. Briefly, $D(\Gamma)^* = \Gamma\backslash D^*$ where D^* is the union of D with certain "rational boundary components" endowed with the Satake topology; again, the automorphic forms of a sufficiently high weight map $\Gamma\backslash D^*$ isomorphically onto a closed subvariety of a projective space, and $\Gamma\backslash D$ is a zariski-open subvariety of $\Gamma\backslash D^*$.

For the Siegel upper half space $\mathcal{H}_g$, the compactification $\mathcal{H}_g^*$ was introduced by Satake (1956) in order to give a geometric foundation to certain results of Siegel (1939), for example, that the space of holomorphic modular forms on $\mathcal{H}_g$ of a fixed weight is finite dimensional, and that the meromorphic functions on $\mathcal{H}_g$ obtained as the quotient of two modular forms of the same weight form an algebraic function field of transcendence degree $g(g+1)/2 = \dim \mathcal{H}_g$ over $\mathbb{C}$.

That the quotient $\Gamma\backslash\mathcal{H}_g^*$ of $\mathcal{H}_g^*$ by an arithmetic group Γ has a projective embedding by modular forms, and hence is a projective variety, was proved in Baily 1958, Cartan 1958, and Satake and Cartan 1958.

The construction of $\mathcal{H}_g^*$ depends on the existence of fundamental domains for the arithmetic group Γ acting on $\mathcal{H}_g$. Weil (1958) used reduction theory to construct fundamental sets (a notion weaker than fundamental domain) for the domains associated with certain classical groups (groups of automorphisms of semsimple $\mathbb{Q}$-algebras with, or without, involution), and Satake (1960) applied this to construct compactifications of these domains. Borel and Harish-Chandra developed a reduction theory for general semisimple groups (Borel and Harish-Chandra 1962; Borel 1962), which then enabled Baily and Borel (1966) to obtain the above theorem in complete generality.

The only source for the proof is the original paper, although some simplifications to the proof are known.

Remark 3.13. (a) The variety $D(\Gamma)^*$ is usually very singular. The boundary $D(\Gamma)^* \smallsetminus D(\Gamma)$ has codimension ≥ 2, provided PGL_2 is not a quotient of the $\mathbb{Q}$-group G giving rise to Γ.

(b) The variety $D(\Gamma)^* = \mathrm{Proj}(\bigoplus_{n\geq 0} A_n)$ where A_n is the vector space of automorphic forms for the n^{th} power of the canonical automorphy factor (Baily and Borel 1966, 10.11). It follows that, if PGL_2 is not a quotient of G, then $D(\Gamma)^* = \mathrm{Proj}(\bigoplus_{n\geq 0} H^0(D(\Gamma), \omega^n))$ where ω is the sheaf of algebraic differentials of maximum degree on $D(\Gamma)$. Without the condition on G, there is a similar description of $D(\Gamma)^*$ in terms of differentials with logarithmic poles (Brylinski 1983, 4.1.4; Mumford 1977).

(b) When $D(\Gamma)$ is compact, Theorem 3.12 follows from the Kodaira embedding theorem (Wells 1980, VI 4.1, 1.5). Nadel and Tsuji (1988, 3.1) extended this to those $D(\Gamma)$ having boundary of dimension 0, and Mok and Zhong (1989) give an alternative proof of Theorem 3.12, but without the information on the boundary given by the original proof.

An algebraic variety $D(\Gamma)$ arising as in the theorem is called a ***locally symmetric variety*** (or an ***arithmetic locally symmetric variety***, or an ***arithmetic variety***, but not yet a Shimura variety).

The theorem of Borel.

THEOREM 3.14 (Borel 1972). *Let $D(\Gamma)$ and $D(\Gamma)^*$ be as in (3.12) — in particular, Γ is torsion free and arithmetic. Let V be a nonsingular quasi-projective variety over $\mathbb{C}$. Then every holomorphic map $f\colon V^{an} \to D(\Gamma)^{an}$ is regular.*

The key step in Borel's proof is the following result:

LEMMA 3.15. *Let $\mathcal{D}_1^\times$ be the punctured disk $\{z \mid 0 < |z| < 1\}$. Then every holomorphic map[8] $\mathcal{D}_1^{\times r} \times \mathcal{D}_1^s \to D(\Gamma)$ extends to a holomorphic map $\mathcal{D}_1^{r+s} \to D(\Gamma)^*$ (of complex spaces).*

The original result of this kind is the big Picard theorem, which, interestingly, was first proved using elliptic modular functions. Recall that the theorem says that if a function f has an essential singularity at a point $p \in \mathbb{C}$, then on any open disk containing p, f takes every complex value except possibly one. Therefore, if a holomorphic function f on $\mathcal{D}_1^\times$ omits two values in $\mathbb{C}$, then it has at worst a pole at 0, and so extends to a holomorphic function $\mathcal{D}_1 \to \mathbb{P}^1(\mathbb{C})$. This can be restated as follows: every holomorphic function from $\mathcal{D}_1^\times$ to $\mathbb{P}^1(\mathbb{C}) \smallsetminus \{3 \text{ points}\}$ extends to a holomorphic function from $\mathcal{D}_1$ to the natural compactification $\mathbb{P}^1(\mathbb{C})$ of $\mathbb{P}^1(\mathbb{C}) \smallsetminus \{3$ points$\}$. Over the decades, there were various improvements made to this theorem. For example, Kwack (1969) replaced $\mathbb{P}^1(\mathbb{C}) \smallsetminus \{3 \text{ points}\}$ with a more general class of spaces. Borel (1972) verified that Kwack's theorem applies to $D(\Gamma) \subset D(\Gamma)^*$, and extended the result to maps from a product $\mathcal{D}_1^{\times r} \times \mathcal{D}_1^s$.

Using the lemma, we can prove the theorem. According Hironaka's (Fields medal) theorem on the resolution of singularities (Hironaka 1964; see also Bravo et al. 2002), we can realize V as an open subvariety of a projective nonsingular variety V^* in such a way that $V^* \smallsetminus V$ is a divisor with normal crossings. This means that, locally for the complex topology, the inclusion $V \hookrightarrow V^*$ is of the form $\mathcal{D}_1^{\times r} \times \mathcal{D}_1^s \hookrightarrow \mathcal{D}_1^{r+s}$. Therefore, the lemma shows that $f\colon V^{\mathrm{an}} \to D(\Gamma)^{\mathrm{an}}$ extends to a holomorphic map $V^{*\mathrm{an}} \to D(\Gamma)^*$, which is regular by Chow's theorem (3.11).

COROLLARY 3.16. *The structure of an algebraic variety on $D(\Gamma)$ is unique.*

PROOF. Let $D(\Gamma)$ denote $\Gamma \backslash D$ with the canonical algebraic structure provided by Theorem 3.12, and suppose $\Gamma \backslash D = V^{\mathrm{an}}$ for a second variety V. Then the identity map $f\colon V^{\mathrm{an}} \to D(\Gamma)$ is a regular bijective map of nonsingular varieties in characteristic zero, and is therefore an isomorphism (cf. AG 8.19). □

The proof of the theorem shows that the compactification $D(\Gamma) \hookrightarrow D(\Gamma)^*$ has the following property: for any compactification $D(\Gamma) \to D(\Gamma)^\dagger$ with $D(\Gamma)^\dagger \smallsetminus D(\Gamma)$ a divisor with normal crossings, there is a unique regular map $D(\Gamma)^\dagger \to D(\Gamma)^*$ making

[8] Recall that $\mathcal{D}_1$ is the open unit disk. The product $\mathcal{D}_1^{\times r} \times \mathcal{D}_1^s$ is obtained from $\mathcal{D}_1^{r+s}$ by removing the first r coordinate hyperplanes.

$$\begin{tikzcd} & D(\Gamma)^{\dagger} \arrow[dd] \\ D(\Gamma) \arrow[ur] \arrow[dr] & \\ & D(\Gamma)^{*} \end{tikzcd}$$

commute. For this reason, $D(\Gamma) \hookrightarrow D(\Gamma)^{*}$ is often called the ***minimal*** compactification. Other names: ***standard***, ***Satake-Baily-Borel***, ***Baily-Borel.***

ASIDE 3.17. (a) Theorem 3.14 also holds for singular V — in fact, it suffices to show that f becomes regular when restricted to an open dense set of V, which we may take to be the complement of the singular locus.

(b) Theorem 3.14 definitely fails without the condition that Γ be torsion free. For example, it is false for $\Gamma\backslash\mathcal{H}_1 = \mathbb{A}^1$ — consider $z \mapsto e^z : \mathbb{C} \to \mathbb{C}$.

Finiteness of the group of automorphisms of $D(\Gamma)$.

DEFINITION 3.18. A semisimple group G over $\mathbb{Q}$ is said to be of ***compact type*** if $G(\mathbb{R})$ is compact, and it is of ***noncompact type*** if it does not contain a nonzero normal subgroup of compact type.

A semisimple group over $\mathbb{Q}$ is an almost direct product of its minimal connected normal subgroups, and it will be of noncompact type if and only if none of these subgroups is of compact type. In particular, a simply connected or adjoint group is of noncompact type if and only if it has no simple factor of compact type.

We shall need one last result about arithmetic subgroups.

THEOREM 3.19 (Borel density theorem). *Let G be a semisimple group over $\mathbb{Q}$ of noncompact type. Then every arithmetic subgroup Γ of $G(\mathbb{Q})$ is zariski-dense in G.*

PROOF. Borel 1969, 15.12, or Platonov and Rapinchuk 1994, Theorem 4.10, p205. □

COROLLARY 3.20. *For G as in (3.19), the centralizer of Γ in $G(\mathbb{R})$ is $Z(\mathbb{R})$, where Z is the centre of G (as an algebraic group over $\mathbb{Q}$).*

PROOF. The theorem implies that the centralizer of Γ in $G(\mathbb{C})$ is $Z(\mathbb{C})$, and $Z(\mathbb{R}) = Z(\mathbb{C}) \cap G(\mathbb{R})$. □

THEOREM 3.21. *Let $D(\Gamma)$ be the quotient of a hermitian symmetric domain D by a torsion free arithmetic group Γ. Then $D(\Gamma)$ has only finitely many automorphisms.*

PROOF. As Γ is a torsion free, D is the universal covering space of $\Gamma\backslash D$ and Γ is the group of covering transformations (see p287). An automorphism $\alpha\colon \Gamma\backslash D \to \Gamma\backslash D$ lifts to an automorphism $\tilde{\alpha}\colon D \to D$. For any $\gamma \in \Gamma$, $\tilde{\alpha}\gamma\tilde{\alpha}^{-1}$ is a covering transformation, and so lies in Γ. Conversely, an automorphism of D normalizing Γ defines an automorphism of $\Gamma\backslash D$. Thus,

$$\mathrm{Aut}(\Gamma\backslash D) = N/\Gamma, \quad N = \text{normalizer of } \Gamma \text{ in } \mathrm{Aut}(D).$$

The corollary implies that the map $\mathrm{ad}\colon N \to \mathrm{Aut}(\Gamma)$ is injective. The group Γ is countable because it is a discrete subgroup of a group that admits a countable basis for its open subsets, and so N is also countable. Because Γ is closed in $\mathrm{Aut}(D)$, so also is N. Write N as a countable union of its finite subsets. According to the Baire category theorem (MF 1.3) one of the finite sets must have an interior point, and this implies that N is discrete. Because $\Gamma\backslash \mathrm{Aut}(D)$ has finite volume (3.3a), this implies that Γ has finite index in N.

Alternatively, there is a geometric proof, at least when Γ is neat. According to Mumford 1977, Proposition 4.2, $D(\Gamma)$ is then an algebraic variety of logarithmic general type, which implies that its automorphism group is finite (Iitaka 1982, 11.12). □

ASIDE 3.22. In most of this section we have considered only quotients $\Gamma\backslash D$ with Γ torsion free. In particular, we disallowed $\Gamma(1)\backslash\mathcal{H}_1$. Typically, if Γ has torsion, then $\Gamma\backslash D$ will be singular and some of the above statements will fail for $\Gamma\backslash D$.

NOTES. Borel 1969, Raghunathan 1972, and (eventually) Witte 2001 contain good expositions on discrete subgroups of Lie groups. There is a large literature on the various compactifications of locally symmetric varieties. For overviews, see Satake 2001 and Goresky 2003, and for a detailed description of the construction of toroidal compactifications, which, in contrast to the Baily-Borel compactification, may be smooth and projective, see Ash et al. 1975.

4. Connected Shimura varieties

Congruence subgroups. Let G be a reductive algebraic group over $\mathbb{Q}$. Choose an embedding $G \hookrightarrow \mathrm{GL}_n$, and define

$$\Gamma(N) = G(\mathbb{Q}) \cap \{g \in \mathrm{GL}_n(\mathbb{Z}) \mid g \equiv I_n \bmod N\}.$$

For example, if $G = \mathrm{SL}_2$, then

$$\Gamma(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \mid ad - bc = 1, \quad a, d \equiv 1, \quad b, c \equiv 0 \mod N \right\}.$$

A ***congruence subgroup of*** $G(\mathbb{Q})$ is any subgroup containing some $\Gamma(N)$ as a subgroup of finite index. Although $\Gamma(N)$ depends on the choice the embedding, this definition does not (see 4.1 below).

With this terminology, a subgroup of $G(\mathbb{Q})$ is arithmetic if it is commensurable with $\Gamma(1)$. The classical congruence subgroup problem for G asks whether every arithmetic subgroup of $G(\mathbb{Q})$ is congruence, i.e., contains some $\Gamma(N)$. For split simply connected groups other than SL_2, the answer is yes (Matsumoto 1969), but SL_2 and all nonsimply connected groups have many noncongruence arithmetic subgroups (for a discussion of the problem, see Platonov and Rapinchuk 1994, section 9.5). In contrast to arithmetic subgroups, the image of a congruence subgroup under an isogeny of algebraic groups need not be a congruence subgroup.

The ***ring of finite adèles*** is the restricted topological product

$$\mathbb{A}_f = \prod (\mathbb{Q}_\ell \colon \mathbb{Z}_\ell)$$

where ℓ runs over the finite primes of ℓ (that is, we omit the factor $\mathbb{R}$). Thus, $\mathbb{A}_f$ is the subring of $\prod \mathbb{Q}_\ell$ consisting of the (a_ℓ) such that $a_\ell \in \mathbb{Z}_\ell$ for almost all ℓ, and it is endowed with the topology for which $\prod \mathbb{Z}_\ell$ is open and has the product topology.

Let $V = \operatorname{Specm} A$ be an affine variety over $\mathbb{Q}$. The set of points of V with coordinates in a $\mathbb{Q}$-algebra R is

$$V(R) = \operatorname{Hom}_{\mathbb{Q}}(A, R).$$

When we write

$$A = \mathbb{Q}[X_1, \ldots, X_m]/\mathfrak{a} = \mathbb{Q}[x_1, \ldots, x_m],$$

the map $P \mapsto (P(x_1), \ldots, P(x_m))$ identifies $V(R)$ with

$$\{(a_1, \ldots, a_m) \in R^m \mid f(a_1, \ldots, a_m) = 0, \quad \forall f \in \mathfrak{a}\}.$$

Let $\mathbb{Z}[x_1, \ldots, x_m]$ be the $\mathbb{Z}$-subalgebra of A generated by the x_i, and let

$$V(\mathbb{Z}_\ell) = \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}[x_1, \ldots, x_m], \mathbb{Z}_\ell) = V(\mathbb{Q}_\ell) \cap \mathbb{Z}_\ell^m \qquad (\text{inside } \mathbb{Q}_\ell^m).$$

This set depends on the choice of the generators x_i for A, but if $A = \mathbb{Q}[y_1, \ldots, y_n]$, then the y_i's can be expressed as polynomials in the x_i with coefficients in $\mathbb{Q}$, and vice versa. For some $d \in \mathbb{Z}$, the coefficients of these polynomials lie in $\mathbb{Z}[\frac{1}{d}]$, and so

$$\mathbb{Z}[\tfrac{1}{d}][x_1, \ldots, x_m] = \mathbb{Z}[\tfrac{1}{d}][y_1, \ldots, y_n] \qquad (\text{inside } A).$$

It follows that for $\ell \nmid d$, the y_i's give the same set $V(\mathbb{Z}_\ell)$ as the x_i's. Therefore,

$$V(\mathbb{A}_f) = \prod (V(\mathbb{Q}_\ell) \colon V(\mathbb{Z}_\ell))$$

is independent of the choice of generators for[9] A.

For an algebraic group G over $\mathbb{Q}$, we define

$$G(\mathbb{A}_f) = \prod (G(\mathbb{Q}_\ell) \colon G(\mathbb{Z}_\ell))$$

similarly. For example,

$$\mathbb{G}_m(\mathbb{A}_f) = \prod (\mathbb{Q}_\ell^\times : \mathbb{Z}_\ell^\times) = \mathbb{A}_f^\times.$$

Proposition 4.1. *For any compact open subgroup K of $G(\mathbb{A}_f)$, $K \cap G(\mathbb{Q})$ is a congruence subgroup of $G(\mathbb{Q})$, and every congruence subgroup arises in this way.*

Proof. Fix an embedding $G \hookrightarrow \mathrm{GL}_n$. From this we get a surjection $\mathbb{Q}[\mathrm{GL}_n] \to \mathbb{Q}[G]$ (of $\mathbb{Q}$-algebras of regular functions), i.e., a surjection

$$\mathbb{Q}[X_{11}, \ldots, X_{nn}, T]/(\det(X_{ij})T - 1) \to \mathbb{Q}[G],$$

and hence $\mathbb{Q}[G] = \mathbb{Q}[x_{11}, \ldots, x_{nn}, t]$. For this presentation of $\mathbb{Q}[G]$,

$$G(\mathbb{Z}_\ell) = G(\mathbb{Q}_\ell) \cap \mathrm{GL}_n(\mathbb{Z}_\ell) \qquad (\text{inside } \mathrm{GL}_n(\mathbb{Q}_\ell)).$$

For an integer $N > 0$, let

$$K(N) = \textstyle\prod_\ell K_\ell, \quad \text{where} \quad K_\ell = \begin{cases} G(\mathbb{Z}_\ell) & \text{if } \ell \nmid N \\ \{g \in G(\mathbb{Z}_\ell) \mid g \equiv I_n \bmod \ell^{r_\ell}\} & \text{if } r_\ell = \operatorname{ord}_\ell(N). \end{cases}$$

Then $K(N)$ is a compact open subgroup of $G(\mathbb{A}_f)$, and

$$K(N) \cap G(\mathbb{Q}) = \Gamma(N).$$

It follows that the compact open subgroups of $G(\mathbb{A}_f)$ containing $K(N)$ intersect $G(\mathbb{Q})$ exactly in the congruence subgroups of $G(\mathbb{Q})$ containing $\Gamma(N)$. Since every

[9] In a more geometric language, let $\alpha \colon V \hookrightarrow \mathbb{A}^m_{\mathbb{Q}}$ be a closed immersion. The zariski closure V_α of V in $\mathbb{A}^m_{\mathbb{Z}}$ is a model of V flat over $\operatorname{Spec}\mathbb{Z}$. A different closed immersion β gives a different flat model V_β, but for some d, the isomorphism $(V_\alpha)_{\mathbb{Q}} \cong V \cong (V_\beta)_{\mathbb{Q}}$ on generic fibres extends to an isomorphism $V_\alpha \to V_\beta$ over $\operatorname{Spec}\mathbb{Z}[\frac{1}{d}]$. For the primes ℓ not dividing d, the subgroups $V_\alpha(\mathbb{Z}_\ell)$ and $V_\beta(\mathbb{Z}_\ell)$ of $V(\mathbb{Q}_\ell)$ will coincide.

compact open subgroup of $G(\mathbb{A}_f)$ contains $K(N)$ for some N, this completes the proof. □

REMARK 4.2. There is a topology on $G(\mathbb{Q})$ for which the congruence subgroups form a fundamental system of neighbourhoods. The proposition shows that this topology coincides with that defined by the diagonal embedding $G(\mathbb{Q}) \subset G(\mathbb{A}_f)$.

EXERCISE 4.3. Show that the image in $\mathrm{PGL}_2(\mathbb{Q})$ of a congruence subgroup in $\mathrm{SL}_2(\mathbb{Q})$ need not be congruence.

Connected Shimura data.

DEFINITION 4.4. A ***connected Shimura datum*** is a pair (G, D) consisting of a semisimple algebraic group G over $\mathbb{Q}$ and a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class D of homomorphisms $u\colon U_1 \to G^{\mathrm{ad}}_{\mathbb{R}}$ satisfying the following conditions:

SU1**:** for $u \in D$, only the characters $z, 1, z^{-1}$ occur in the representation of U_1 on $\mathrm{Lie}(G^{\mathrm{ad}})_{\mathbb{C}}$ defined by u;

SU2**:** for $u \in D$, $\mathrm{ad}u(-1)$ is a Cartan involution on G^{ad};

SU3**:** G^{ad} has no $\mathbb{Q}$-factor H such that $H(\mathbb{R})$ is compact.

EXAMPLE 4.5. Let $u\colon U_1 \to \mathrm{PGL}_2(\mathbb{R})$ be the homomorphism sending $z = (a+bi)^2$ to $\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) \bmod \pm I_2$ (cf. 1.10), and let D be the set of conjugates of this homomorphism, i.e., D is the set of homomorphisms $U_1 \to \mathrm{PGL}_2(\mathbb{R})$ of the form

$$z = (a+bi)^2 \mapsto A \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) A^{-1} \bmod \pm I_2, \quad A \in \mathrm{SL}_2(\mathbb{R}).$$

Then (SL_2, D) is a Shimura datum (here SL_2 is regarded as a group over $\mathbb{Q}$).

REMARK 4.6. (a) If $u\colon U_1 \to G^{\mathrm{ad}}(\mathbb{R})$ satisfies the conditions SU1,2, then so does any conjugate of it by an element of $G^{\mathrm{ad}}(\mathbb{R})^+$. Thus a pair (G, u) satisfying SU1,2,3 determines a connected Shimura datum. Our definition of connected Shimura datum was phrased so as to avoid D having a distinguished point.

(b) Condition SU3 says that G is of noncompact type (3.18). It is fairly harmless to assume this, because replacing G with its quotient by a connected normal subgroup N such that $N(\mathbb{R})$ is compact changes little. Assuming it allows us to apply the strong approximation theorem when G is simply connected (see 4.16 below).

LEMMA 4.7. *Let H be an adjoint real Lie group, and let $u\colon U_1 \to H$ be a homomorphism satisfying SU1,2. Then the following conditions on u are equivalent:*

(a) $u(-1) = 1$;
(b) *u is trivial, i.e., $u(z) = 1$ for all z;*
(c) *H is compact.*

PROOF. (a)⇔(b). If $u(-1) = 1$, then u factors through $U_1 \xrightarrow{2} U_1$, and so $z^{\pm 1}$ can not occur in the representation of U_1 on $\mathrm{Lie}(H)_{\mathbb{C}}$. Therefore U_1 acts trivially on $\mathrm{Lie}(H)_{\mathbb{C}}$, which implies (b). The converse is trivial.

(a)⇔(c). We have

$$H \text{ is compact} \overset{1.17a}{\iff} \mathrm{ad}u(-1) = 1 \overset{Z(H)=1}{\iff} u(-1) = 1. \quad \square$$

PROPOSITION 4.8. *To give a connected Shimura datum is the same as to give*

- *a semisimple algebraic group G over $\mathbb{Q}$ of noncompact type,*
- *a hermitian symmetric domain D, and*

- *an action of $G(\mathbb{R})^+$ on D defined by a surjective homomorphism $G(\mathbb{R})^+ \to \mathrm{Hol}(D)^+$ with compact kernel.*

PROOF. Let (G, D) be a connected Shimura datum, and let $u \in D$. Decompose $G^{\mathrm{ad}}_{\mathbb{R}}$ into a product of its simple factors: $G^{\mathrm{ad}}_{\mathbb{R}} = H_1 \times \cdots \times H_s$. Correspondingly, $u = (u_1, \ldots, u_s)$ where u_i is the projection of u into $H_i(\mathbb{R})$. Then $u_i = 1$ if H_i is compact (4.7), and otherwise there is an irreducible hermitian symmetric domain D'_i such that $H_i(\mathbb{R})^+ = \mathrm{Hol}(D'_i)^+$ and D'_i is in natural one-to-one correspondence with the set D_i of $H_i(\mathbb{R})^+$-conjugates of u_i (see 1.21). The product D' of the D'_i is a hermitian symmetric domain on which $G(\mathbb{R})^+$ acts via a surjective homomorphism $G(\mathbb{R})^+ \to \mathrm{Hol}(D)^+$ with compact kernel. Moreover, there is a natural identification of $D' = \prod D'_i$ with $D = \prod D_i$.

Conversely, let $(G, D, G(\mathbb{R})^+ \to \mathrm{Hol}(D)^+)$ satisfy the conditions in the proposition. Decompose $G^{\mathrm{ad}}_{\mathbb{R}}$ as before, and let H_{c} (resp. H_{nc}) be the product of the compact (resp. noncompact) factors. The action of $G(\mathbb{R})^+$ on D defines an isomorphism $H_{\mathrm{nc}}(\mathbb{R})^+ \cong \mathrm{Hol}(D)^+$, and $\{u_p \mid p \in D\}$ is an $H_{\mathrm{nc}}(\mathbb{R})^+$-conjugacy class of homomorphisms $U_1 \to H_{\mathrm{nc}}(\mathbb{R})^+$ satisfying SU1,2 (see 1.21). Now

$$\{(1, u_p)\colon U_1 \to H_{\mathrm{c}}(\mathbb{R}) \times H_{\mathrm{nc}}(\mathbb{R}) \mid p \in D\},$$

is a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class of homomorphisms $U_1 \to G^{\mathrm{ad}}(\mathbb{R})$ satisfying SU1,2. □

PROPOSITION 4.9. *Let (G, D) be a connected Shimura datum, and let X be the $G^{\mathrm{ad}}(\mathbb{R})$-conjugacy class of homomorphisms $\mathbb{S} \to G_{\mathbb{R}}$ containing D. Then D is a connected component of X, and the stabilizer of D in $G^{\mathrm{ad}}(\mathbb{R})$ is $G^{\mathrm{ad}}(\mathbb{R})^+$.*

PROOF. The argument in the proof of (1.5) shows that X is a disjoint union of orbits $G^{\mathrm{ad}}(\mathbb{R})^+ h$, each of which is both open and closed in X. In particular, D is a connected component of X.

Let H_{c} (resp. H_{nc}) be the product of the compact (resp. noncompact) simple factors of $G_{\mathbb{R}}$. Then H_{nc} is a connected algebraic group over $\mathbb{R}$ such that $H_{\mathrm{nc}}(\mathbb{R})^+ = \mathrm{Hol}(D)$, and $G(\mathbb{R})^+$ acts on D through its quotient $H_{\mathrm{nc}}(\mathbb{R})^+$. As $H_{\mathrm{c}}(\mathbb{R})$ is connected (Borel 1991, p277), the last part of the proposition follows from (1.7). □

Definition of a connected Shimura variety. Let (G, D) be a connected Shimura datum, and regard D as a hermitian symmetric domain with $G(\mathbb{R})^+$ acting on it as in (4.8). Because $G^{\mathrm{ad}}(\mathbb{R})^+ \to \mathrm{Aut}(D)^+$ has compact kernel, the image $\overline{\Gamma}$ of any arithmetic subgroup Γ of $G^{\mathrm{ad}}(\mathbb{Q})^+$ in $\mathrm{Aut}(D)^+$ will be arithmetic (this is the definition p289). The kernel of $\Gamma \to \overline{\Gamma}$ is finite. If Γ is torsion free, then $\Gamma \cong \overline{\Gamma}$, and so the Baily-Borel and Borel theorems (3.12, 3.14) apply to

$$D(\Gamma) \overset{\mathrm{df}}{=} \Gamma \backslash D = \overline{\Gamma} \backslash D.$$

In particular, $D(\Gamma)$ is an algebraic variety, and, for any $\Gamma \supset \Gamma'$, the natural map

$$D(\Gamma) \leftarrow D(\Gamma')$$

is regular.

DEFINITION 4.10. The ***connected Shimura variety*** $\mathrm{Sh}^{\circ}(G, D)$ is the inverse system of locally symmetric varieties $(D(\Gamma))_{\Gamma}$ where Γ runs over the torsion-free arithmetic subgroups of $G^{\mathrm{ad}}(\mathbb{Q})^+$ whose inverse image in $G(\mathbb{Q})^+$ is a congruence subgroup.

REMARK 4.11. An element g of $G^{\mathrm{ad}}(\mathbb{Q})^+$ defines a holomorphic map $g\colon D \to D$, and hence a map

$$\Gamma\backslash D \to g\Gamma g^{-1}\backslash D.$$

This is again holomorphic (3.1), and hence is regular (3.14). Therefore the group $G^{\mathrm{ad}}(\mathbb{Q})^+$ acts on the family $\mathrm{Sh}^\circ(G,D)$ (but not on the individual $D(\Gamma)$'s).

LEMMA 4.12. *Write π for the homomorphism $G(\mathbb{Q})^+ \to G^{\mathrm{ad}}(\mathbb{Q})^+$. The following conditions on an arithmetic subgroup Γ of $G^{\mathrm{ad}}(\mathbb{Q})^+$ are equivalent:*

(a) *$\pi^{-1}(\Gamma)$ is a congruence subgroup of $G(\mathbb{Q})^+$;*
(b) *$\pi^{-1}(\Gamma)$ contains a congruence subgroup of $G(\mathbb{Q})^+$;*
(c) *Γ contains the image of a congruence subgroup of $G(\mathbb{Q})^+$.*

Therefore, the varieties $\Gamma\backslash D$ with Γ a congruence subgroup of $G(\mathbb{Q})^+$ such $\pi(\Gamma)$ is torsion free are cofinal in the family $\mathrm{Sh}^\circ(G,D)$.

PROOF. (a) $\Longrightarrow$ (b). Obvious.

(b) $\Longrightarrow$ (c). Let Γ' be a congruence subgroup of $G(\mathbb{Q})^+$ contained in $\pi^{-1}(\Gamma)$. Then

$$\Gamma \supset \pi(\pi^{-1}(\Gamma)) \supset \pi(\Gamma').$$

(c) $\Longrightarrow$ (a). Let Γ' be a congruence subgroup of $G(\mathbb{Q})^+$ such that $\Gamma \supset \pi(\Gamma')$, and consider

$$\pi^{-1}(\Gamma) \supset \pi^{-1}\pi(\Gamma') \supset \Gamma'.$$

Because $\pi(\Gamma')$ is arithmetic (3.2), it is of finite index in Γ, and it follows that $\pi^{-1}\pi(\Gamma')$ is of finite index in $\pi^{-1}(\Gamma)$. Because $Z(\mathbb{Q})\cdot\Gamma' \supset \pi^{-1}\pi(\Gamma')$ and $Z(\mathbb{Q})$ is finite (Z is the centre of G), Γ' is of finite index in $\pi^{-1}\pi(\Gamma')$. Therefore, Γ' is of finite index in $\pi^{-1}(\Gamma)$, which proves that $\pi^{-1}(\Gamma)$ is congruence. □

REMARK 4.13. The homomorphism $\pi\colon G(\mathbb{Q})^+ \to G^{\mathrm{ad}}(\mathbb{Q})^+$ is usually far from surjective. Therefore, $\pi\pi^{-1}(\Gamma)$ is usually not equal to Γ, and the family $D(\Gamma)$ with Γ a congruence subgroup of $G(\mathbb{Q})^+$ is usually much smaller than $\mathrm{Sh}^\circ(G,D)$.

EXAMPLE 4.14. (a) $G = \mathrm{SL}_2$, $D = \mathcal{H}_1$. Then $\mathrm{Sh}^\circ(G,D)$ is the family of elliptic modular curves $\Gamma\backslash\mathcal{H}_1$ with Γ a torsion-free arithmetic subgroup of $\mathrm{PGL}_2(\mathbb{R})^+$ containing the image of $\Gamma(N)$ for some N.

(b) $G = \mathrm{PGL}_2$, $D = \mathcal{H}_1$. The same as (a), except that now the Γ are required to be congruence subgroups of $\mathrm{PGL}_2(\mathbb{Q})$ — there are *many fewer* of these (see 4.3).

(c) Let B be a quaternion algebra over a totally real field F. Then

$$B \otimes_{\mathbb{Q}} \mathbb{R} \cong \textstyle\prod_{v\colon F\hookrightarrow\mathbb{R}} B \otimes_{F,v} \mathbb{R}$$

and each $B \otimes_{F,v} \mathbb{R}$ is isomorphic either to the usual quaternions $\mathbb{H}$ or to $M_2(\mathbb{R})$. Let G be the semisimple algebraic group over $\mathbb{Q}$ such that

$$G(\mathbb{Q}) = \mathrm{Ker}(\mathrm{Nm}\colon B^\times \to F^\times).$$

Then

$$G(\mathbb{R}) \approx \mathbb{H}^{\times 1} \times \cdots \times \mathbb{H}^{\times 1} \times \mathrm{SL}_2(\mathbb{R}) \times \cdots \times \mathrm{SL}_2(\mathbb{R}) \tag{27}$$

where $\mathbb{H}^{\times 1} = \mathrm{Ker}(\mathrm{Nm}\colon \mathbb{H}^\times \to \mathbb{R}^\times)$. Assume that at least one $\mathrm{SL}_2(\mathbb{R})$ occurs (so that G is of noncompact type), and let D be a product of copies of $\mathcal{H}_1$, one for each copy of $\mathrm{SL}_2(\mathbb{R})$. The choice of an isomorphism (27) determines an action of $G(\mathbb{R})$ on D which satisfies the conditions of (4.8), and hence defines a connected Shimura datum. In this case, $D(\Gamma)$ has dimension equal to the number of copies of $M_2(\mathbb{R})$

in the decomposition of $B \otimes_{\mathbb{Q}} \mathbb{R}$. If $B \approx M_2(F)$, then $G(\mathbb{Q})$ has unipotent elements, e.g., $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, and so $D(\Gamma)$ is not compact (3.3). In this case the varieties $D(\Gamma)$ are called ***Hilbert modular varieties***. On the other hand, if B is a division algebra, $G(\mathbb{Q})$ has no unipotent elements, and so the $D(\Gamma)$ are compact (as manifolds, hence they are projective as algebraic varieties).

ASIDE 4.15. In the definition of $\mathrm{Sh}^{\circ}(G, D)$, why do we require the inverse images of the Γ's in $G(\mathbb{Q})^{+}$ to be congruence? The arithmetic properties of the quotients of hermitian symmetric domains by noncongruence arithmetic subgroups are not well understood even for $D = \mathcal{H}_1$ and $G = \mathrm{SL}_2$. Also, the congruence subgroups turn up naturally when we work adèlically.

The strong approximation theorem. Recall that a semisimple group G is said to be simply connected if any isogeny $G' \to G$ with G' connected is an isomorphism. For example, SL_2 is simply connected, but PGL_2 is not.

THEOREM 4.16 (Strong Approximation). *Let G be an algebraic group over $\mathbb{Q}$. If G is semisimple, simply connected, and of noncompact type, then $G(\mathbb{Q})$ is dense in $G(\mathbb{A}_f)$.*

PROOF. Platonov and Rapinchuk 1994, Theorem 7.12, p427. □

REMARK 4.17. Without the conditions on G, the theorem fails, as the following examples illustrate:

(a) $\mathbb{G}_m$: the group $\mathbb{Q}^{\times}$ is not dense in $\mathbb{A}_f^{\times}$.

(b) PGL_2: the determinant defines surjections

$$\mathrm{PGL}_2(\mathbb{Q}) \to \mathbb{Q}^{\times}/\mathbb{Q}^{\times 2}$$
$$\mathrm{PGL}_2(\mathbb{A}_f) \to \mathbb{A}_f^{\times}/\mathbb{A}_f^{\times 2}$$

and $\mathbb{Q}^{\times}/\mathbb{Q}^{\times 2}$ is not dense in $\mathbb{A}_f^{\times}/\mathbb{A}_f^{\times 2}$.

(c) G of compact type: because $G(\mathbb{Z})$ is discrete in $G(\mathbb{R})$ (see 3.3), it is finite, and so it is not dense in $G(\hat{\mathbb{Z}})$, which implies that $G(\mathbb{Q})$ is not dense in $G(\mathbb{A}_f)$.

An adèlic description of $D(\Gamma)$.

PROPOSITION 4.18. *Let (G, D) be a connected Shimura datum with G simply connected. Let K be a compact open subgroup of $G(\mathbb{A}_f)$, and let*

$$\Gamma = K \cap G(\mathbb{Q})$$

be the corresponding congruence subgroup of $G(\mathbb{Q})$. The map $x \mapsto [x, 1]$ defines a bijection

$$\Gamma \backslash D \cong G(\mathbb{Q}) \backslash D \times G(\mathbb{A}_f)/K. \tag{28}$$

Here $G(\mathbb{Q})$ acts on both D and $G(\mathbb{A}_f)$ on the left, and K acts on $G(\mathbb{A}_f)$ on the right:

$$q \cdot (x, a) \cdot k = (qx, qak), \quad q \in G(\mathbb{Q}), \quad x \in D, \quad a \in G(\mathbb{A}_f), \quad k \in K.$$

When we endow D with its usual topology and $G(\mathbb{A}_f)$ with the adèlic topology (or the discrete topology), this becomes a homeomorphism.

PROOF. Because K is open, $G(\mathbb{A}_f) = G(\mathbb{Q}) \cdot K$ (strong approximation theorem). Therefore, every element of $G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)/K$ is represented by an element of the form $[x, 1]$. By definition, $[x, 1] = [x', 1]$ if and only if there exist $q \in G(\mathbb{Q})$ and $k \in K$ such that $x' = qx$, $1 = qk$. The second equation implies that $q = k^{-1} \in \Gamma$, and so $[x, 1] = [x', 1]$ if and only if x and x' represent the same element in $\Gamma\backslash D$.

Consider

$$\begin{array}{ccc} D & \xrightarrow{x\mapsto(x,[1])} & D \times (G(\mathbb{A}_f)/K) \\ \downarrow & & \downarrow \\ \Gamma\backslash D & \xrightarrow{[x]\mapsto[x,1]} & G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)/K. \end{array}$$

As K is open, $G(\mathbb{A}_f)/K$ is discrete, and so the upper map is a homeomorphism of D onto its image, which is open. It follows easily that the lower map is a homeomorphism. □

What happens when we pass to the inverse limit over Γ? The obvious map

$$D \to \varprojlim \Gamma\backslash D,$$

is injective because each Γ acts freely on D and $\bigcap \Gamma = \{1\}$. Is the map surjective? The example

$$\mathbb{Z} \to \varprojlim \mathbb{Z}/m\mathbb{Z} = \hat{\mathbb{Z}}$$

is not encouraging — it suggests that $\varprojlim \Gamma\backslash D$ might be some sort of completion of D relative to the Γ's. This is correct: $\varprojlim \Gamma\backslash D$ is much larger than D. In fact, when we pass to the limit on the right in (28), we get the obvious answer:

PROPOSITION 4.19. *In the limit,*

$$\varprojlim{}_K G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)/K = G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f) \tag{29}$$

(adèlic topology on $G(\mathbb{A}_f)$).

Before proving this, we need a lemma.

LEMMA 4.20. *Let G be a topological group acting continuously on a topological space X, and let $(G_i)_{i\in I}$ be a directed family of subgroups of G. The canonical map $X/\bigcap G_i \to \varprojlim X/G_i$ is injective if the G_i are compact, and it is surjective if in addition the orbits of the G_i in X are separated.*

PROOF. We shall use that a directed intersection of nonempty compact sets is nonempty, which has the consequence that a directed inverse limit of nonempty compact sets is nonempty.

Assume that each G_i is compact, and let $x, x' \in X$. For each i, let

$$G_i(x, x') = \{g \in G_i \mid xg = x'\}.$$

If x and x' have the same image in $\varprojlim X/G_i$, then the $G_i(x, x')$ are all nonempty. Since each is compact, their intersection is nonempty. For any g in the intersection, $xg = x'$, which shows that x and x' have the same image in $X/\bigcap G_i$.

Now assume that each orbit is separated and hence compact. For any $(x_iG_i)_{i\in I} \in \varprojlim X/G_i$, $\varprojlim x_iG_i$ is nonempty. If $x \in \varprojlim x_iG_i$, then $x \cdot \bigcap G_i$ maps to $(x_iG_i)_{i\in I}$. □

PROOF OF 4.19. Let $(x, a) \in D \times G(\mathbb{A}_f)$, and let K be a compact open subgroup of $G(\mathbb{A}_f)$. In order to be able to apply the lemma, we have to show that the image of the orbit $(x, a)K$ in $G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)$ is separated for K sufficiently small. Let $\Gamma = G(\mathbb{Q}) \cap aKa^{-1}$ — we may assume that Γ is torsion free (3.5). There exists an open neighbourhood V of x such that $gV \cap V = \emptyset$ for all $g \in \Gamma \smallsetminus \{1\}$ (see the proof of 3.1). For any $(x, b) \in (x, a)K$, $g(V \times aK) \cap (V \times bK) = \emptyset$ for all $g \in G(\mathbb{Q}) \smallsetminus \{1\}$, and so the images of $V \times Ka$ and $V \times Kb$ in $G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)$ separate (x, a) and (x, b). □

ASIDE 4.21. (a) Why replace the single coset space on the left of (28) with the more complicated double coset space on the right? One reason is that it makes transparent that (in this case) there is an action of $G(\mathbb{A}_f)$ on the inverse system $(\Gamma\backslash D)_\Gamma$, and hence, for example, on

$$\varinjlim H^i(\Gamma\backslash D, \mathbb{Q}).$$

Another reason will be seen presently — we use double cosets to define Shimura varieties. Double coset spaces are pervasive in work on the Langlands program.

(b) The inverse limit of the $D(\Gamma)$ exists as a scheme — it is even locally noetherian and regular (cf. 5.30 below).

Alternative definition of connected Shimura data. Recall that $\mathbb{S}$ is the real torus such that $\mathbb{S}(\mathbb{R}) = \mathbb{C}^\times$. The exact sequence

$$0 \to \mathbb{R}^\times \xrightarrow{r \mapsto r^{-1}} \mathbb{C}^\times \xrightarrow{z \mapsto z/\bar{z}} U_1 \to 0$$

arises from an exact sequence of tori

$$0 \to \mathbb{G}_m \xrightarrow{w} \mathbb{S} \longrightarrow U_1 \to 0.$$

Let H be a semisimple real algebraic group with trivial centre. A homomorphism $u \colon U_1 \to H$ defines a homomorphism $h \colon \mathbb{S} \to H$ by the rule $h(z) = u(z/\bar{z})$, and U_1 will act on $\mathrm{Lie}(H)_\mathbb{C}$ through the characters $z, 1, z^{-1}$ if and only if $\mathbb{S}$ acts on $\mathrm{Lie}(H)_\mathbb{C}$ through the characters $z/\bar{z}, 1, \bar{z}/z$. Conversely, let h be a homomorphism $\mathbb{S} \to H$ for which $\mathbb{S}$ acts on $\mathrm{Lie}(H)_\mathbb{C}$ through the characters $z/\bar{z}, 1, \bar{z}/z$. Then $w(\mathbb{G}_m)$ acts trivially on $\mathrm{Lie}(H)_\mathbb{C}$, which implies that h is trivial on $w(\mathbb{G}_m)$ because the adjoint representation $H \to \mathrm{Lie}(H)$ is faithful. Thus, h arises from a u.

Now let G be a semisimple algebraic group over $\mathbb{Q}$. From the above remark, we see that to give a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class D of homomorphisms $u \colon U_1 \to G^{\mathrm{ad}}_\mathbb{R}$ satisfying SU1,2 is the same as to give a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class X^+ of homomorphisms $h \colon \mathbb{S} \to G^{\mathrm{ad}}_\mathbb{R}$ satisfying the following conditions:

SV1**:** for $h \in X^+$, only the characters $z/\bar{z}, 1, \bar{z}/z$ occur in the representation of $\mathbb{S}$ on $\mathrm{Lie}(G^{\mathrm{ad}})_\mathbb{C}$ defined by h;
SV2**:** $\mathrm{ad}h(i)$ is a Cartan involution on G^{ad}.

DEFINITION 4.22. A ***connected Shimura datum*** is a pair (G, X^+) consisting of a semisimple algebraic group over $\mathbb{Q}$ and a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class of homomorphisms $h \colon \mathbb{S} \to G^{\mathrm{ad}}_\mathbb{R}$ satisfying SV1, SV2, and

SV3**:** G^{ad} has no $\mathbb{Q}$-factor on which the projection of h is trivial.

In the presence of the other conditions, SV3 is equivalent to SU3 (see 4.7). Thus, because of the correspondence $u \leftrightarrow h$, this is essentially the same as Definition 4.4.

Definition 4.4 is more convenient when working with only connected Shimura varieties, while Definition 4.22 is more convenient when working with both connected and nonconnected Shimura varieties.

NOTES. Connected Shimura varieties were defined *en passant* in Deligne 1979, 2.1.8.

5. Shimura varieties

Connected Shimura varieties are very natural objects, so why do we need anything more complicated? There are two main reasons. From the perspective of the Langlands program, we should be working with reductive groups, not semisimple groups. More fundamentally, the varieties $D(\Gamma)$ making up a connected Shimura variety $\mathrm{Sh}^\circ(G, D)$ have models over number fields, but the models depend a realization of G as the derived group of a reductive group. Moreover, the number field depends on Γ — as Γ shrinks the field grows. For example, the modular curve $\Gamma(N)\backslash\mathcal{H}_1$ is naturally defined over $\mathbb{Q}[\zeta_N]$, $\zeta_N = e^{2\pi i/N}$. Clearly, for a canonical model we would like all the varieties in the family to be defined over the same field.[10]

How can we do this? Consider the line $Y+i=0$. This is naturally defined over $\mathbb{Q}[i]$, not $\mathbb{Q}$. On the other hand, the variety $Y^2+1=0$ is naturally defined over $\mathbb{Q}$, and over $\mathbb{C}$ it decomposes into a disjoint pair of conjugate lines $(Y-i)(Y+i)=0$. So we have managed to get our variety defined over $\mathbb{Q}$ at the cost of adding other connected components. It is always possible to lower the field of definition of a variety by taking the disjoint union of it with its conjugates. Shimura varieties give a systematic way of doing this for connected Shimura varieties.

Notations for reductive groups. Let G be a reductive group over $\mathbb{Q}$, and let $G \xrightarrow{\mathrm{ad}} G^{\mathrm{ad}}$ be the quotient of G by its centre Z. We let $G(\mathbb{R})_+$ denote the group of elements of $G(\mathbb{R})$ whose image in $G^{\mathrm{ad}}(\mathbb{R})$ lies in its identity component $G^{\mathrm{ad}}(\mathbb{R})^+$, and we let $G(\mathbb{Q})_+ = G(\mathbb{Q}) \cap G(\mathbb{R})_+$. For example, $\mathrm{GL}_2(\mathbb{Q})_+$ consists of the 2×2 matrices with rational coefficients having positive determinant.

For a reductive group G (resp. for GL_n), there are exact sequences

$$
\begin{array}{ll}
1 \longrightarrow G^{\mathrm{der}} \longrightarrow G \xrightarrow{\nu} T \longrightarrow 1 & 1 \longrightarrow \mathrm{SL}_n \longrightarrow \mathrm{GL}_n \xrightarrow{\det} \mathbb{G}_m \longrightarrow 1 \\
1 \longrightarrow Z \longrightarrow G \xrightarrow{\mathrm{ad}} G^{\mathrm{ad}} \longrightarrow 1 & 1 \longrightarrow \mathbb{G}_m \longrightarrow \mathrm{GL}_n \xrightarrow{\mathrm{ad}} \mathrm{PGL}_n \longrightarrow 1 \\
1 \longrightarrow Z' \longrightarrow Z \longrightarrow T \longrightarrow 1 & 1 \longrightarrow \mu_n \longrightarrow \mathbb{G}_m \xrightarrow{x \mapsto x^n} \mathbb{G}_m \longrightarrow 1
\end{array}
$$

Here T (a torus) is the largest commutative quotient of G, and $Z' =_{\mathrm{df}} Z \cap G^{\mathrm{der}}$ (a finite algebraic group) is the centre of G^{der}.

The real points of algebraic groups.

PROPOSITION 5.1. *For a surjective homomorphism $\varphi\colon G \to H$ of algebraic groups over $\mathbb{R}$, $G(\mathbb{R})^+ \to H(\mathbb{R})^+$ is surjective.*

[10] In fact, Shimura has an elegant way of describing a canonical model in which the varieties in the family are defined over different fields, but this doesn't invalidate my statement. Incidentally, Shimura also requires a reductive (not a semisimple) group in order to have a canonical model over a number field. For an explanation of Shimura's point of view in the language of these notes, see Milne and Shih 1981.

Proof. The map $\varphi(\mathbb{R})\colon G(\mathbb{R})^+ \to H(\mathbb{R})^+$ can be regarded as a smooth map of smooth manifolds. As φ is surjective on the tangent spaces at 1, the image of $\varphi(\mathbb{R})$ contains an open neighbourhood of 1 (Boothby 1975, II 7.1). This implies that the image itself is open because it is a group. It is therefore also closed, and this implies that it equals $H(\mathbb{R})^+$. □

Note that $G(\mathbb{R}) \to H(\mathbb{R})$ need not be surjective. For example, $\mathbb{G}_m \xrightarrow{x \mapsto x^n} \mathbb{G}_m$ is surjective as a map of algebraic groups, but the image of $\mathbb{G}_m(\mathbb{R}) \xrightarrow{n} \mathbb{G}_m(\mathbb{R})$ is $\mathbb{G}_m(\mathbb{R})^+$ or $\mathbb{G}_m(\mathbb{R})$ according as n is even or odd. Also $\mathrm{SL}_2 \to \mathrm{PGL}_2$ is surjective, but the image of $\mathrm{SL}_2(\mathbb{R}) \to \mathrm{PGL}_2(\mathbb{R})$ is $\mathrm{PGL}_2(\mathbb{R})^+$.

For a simply connected algebraic group G, $G(\mathbb{C})$ is simply connected as a topological space, but $G(\mathbb{R})$ need not be. For example, $\mathrm{SL}_2(\mathbb{R})$ is not simply connected.

Theorem 5.2 (Cartan 1927). *For a simply connected group G over $\mathbb{R}$, $G(\mathbb{R})$ is connected.*

Proof. See Platonov and Rapinchuk 1994, Theorem 7.6, p407. □

Corollary 5.3. *For a reductive group G over $\mathbb{R}$, $G(\mathbb{R})$ has only finitely many connected components (for the real topology).*[11]

Proof. Because of (5.1), an exact sequence of real algebraic groups

$$1 \to N \to G' \to G \to 1 \tag{30}$$

with $N \subset Z(G')$ gives rise to an exact sequence

$$\pi_0(G'(\mathbb{R})) \to \pi_0(G(\mathbb{R})) \to H^1(\mathbb{R}, N).$$

Let $\tilde{G}$ be the universal covering group of G^{der}. As G is an almost direct product of $Z = Z(G)$ and G^{der}, there is an exact sequence (30) with $G' = Z \times \tilde{G}$ and N finite. Now

- $\pi_0(\tilde{G}(\mathbb{R})) = 0$ because $\tilde{G}$ is simply connected,
- $\pi_0(Z(\mathbb{R}))$ is finite because Z° has finite index in Z and Z° is a quotient (by a finite group) of a product of copies of U_1 and $\mathbb{G}_m$, and
- $H^1(\mathbb{R}, N)$ is finite because N is finite. □

For example, $\mathbb{G}_m^d(\mathbb{R}) = (\mathbb{R}^\times)^d$ has 2^d connected components, and each of $\mathrm{PGL}_2(\mathbb{R})$ and $\mathrm{GL}_2(\mathbb{R})$ has 2 connected components.

Theorem 5.4 (real approximation). *For any connected algebraic group G over $\mathbb{Q}$, $G(\mathbb{Q})$ is dense in $G(\mathbb{R})$.*

Proof. See Platonov and Rapinchuk 1994, Theorem 7.7, p415. □

Shimura data.

Definition 5.5. A ***Shimura datum*** is a pair (G, X) consisting of a reductive group G over $\mathbb{Q}$ and a $G(\mathbb{R})$-conjugacy class X of homomorphisms $h\colon \mathbb{S} \to G_{\mathbb{R}}$ satisfying the conditions SV1, SV2, and SV3 (see p302).

[11] This also follows from the theorem of Whitney 1957: for an algebraic variety V over $\mathbb{R}$, $V(\mathbb{R})$ has only finitely many connected components (for the real topology) — see Platonov and Rapinchuk 1994, Theorem 3.6, p119.

Note that, in contrast to a connected Shimura datum, G is reductive (not semisimple), the homomorphisms h have target $G_{\mathbb{R}}$ (not $G_{\mathbb{R}}^{\mathrm{ad}}$), and X is the full $G(\mathbb{R})$-conjugacy class (not a connected component).

EXAMPLE 5.6. Let $G = \mathrm{GL}_2$ (over $\mathbb{Q}$) and let X be the set of $\mathrm{GL}_2(\mathbb{R})$-conjugates of the homomorphism $h_o\colon \mathbb{S} \to \mathrm{GL}_{2\mathbb{R}}$, $h_o(a+ib) = \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$. Then (G, X) is a Shimura datum. Note that there is a natural bijection $X \to \mathbb{C} \smallsetminus \mathbb{R}$, namely, $h_o \mapsto i$ and $gh_og^{-1} \mapsto gi$. More intrinsically, $h \leftrightarrow z$ if and only if $h(\mathbb{C}^\times)$ is the stabilizer of z in $\mathrm{GL}_2(\mathbb{R})$ and $h(z)$ acts on the tangent space at z as multiplication by $z/\overline{z}$ (rather than $\overline{z}/z$).

PROPOSITION 5.7. *Let G be a reductive group over $\mathbb{R}$. For a homomorphism $h\colon \mathbb{S} \to G$, let $\overline{h}$ be the composite of h with $G \to G^{\mathrm{ad}}$. Let X be a $G(\mathbb{R})$-conjugacy class of homomorphisms $\mathbb{S} \to G$, and let $\overline{X}$ be the $G^{\mathrm{ad}}(\mathbb{R})$-conjugacy class of homomorphisms $\mathbb{S} \to G^{\mathrm{ad}}$ containing the $\overline{h}$ for $h \in X$.*

(a) *The map $h \mapsto \overline{h}\colon X \to \overline{X}$ is injective and its image is a union of connected components of $\overline{X}$.*

(b) *Let X^+ be a connected component of X, and let $\overline{X}^+$ be its image in $\overline{X}$. If (G, X) satisfies the axioms SV1–3 then $(G^{\mathrm{der}}, \overline{X}^+)$ satisfies the axioms SV1–3; moreover, the stabilizer of X^+ in $G(\mathbb{R})$ is $G(\mathbb{R})_+$ (i.e., $gX^+ = X^+ \iff g \in G(\mathbb{R})_+$).*

PROOF. (a) A homomorphism $h\colon \mathbb{S} \to G$ is determined by its projections to T and G^{ad}, because any other homomorphism with the same projections will be of the form he for some regular map $e\colon \mathbb{S} \to Z'$ and e is trivial because $\mathbb{S}$ is connected and Z' is finite. The elements of X all have the same projection to T, because T is commutative, which proves that $h \mapsto \overline{h}\colon X \to \overline{X}$ is injective. For the second part of the statement, use that $G^{\mathrm{ad}}(\mathbb{R})^+$ acts transitively on each connected component of $\overline{X}$ (see 1.5) and $G(\mathbb{R})^+ \to G^{\mathrm{ad}}(\mathbb{R})^+$ is surjective.

(b) The first assertion is obvious. In (a) we showed that $\pi_0(X) \subset \pi_0(\overline{X})$. The stabilizer in $G^{\mathrm{ad}}(\mathbb{R})$ of $[\overline{X}^+]$ is $G^{\mathrm{ad}}(\mathbb{R})^+$ (see 4.9), and so its stabilizer in $G(\mathbb{R})$ is the inverse image of $G^{\mathrm{ad}}(\mathbb{R})^+$ in $G(\mathbb{R})$. $\square$

COROLLARY 5.8. *Let (G, X) be a Shimura datum, and let X^+ be a connected component of X regarded as a $G(\mathbb{R})^+$-conjugacy class of homomorphisms $\mathbb{S} \to G_{\mathbb{R}}^{\mathrm{ad}}$ (5.7). Then (G^{der}, X^+) is a connected Shimura datum. In particular, X is a finite disjoint union of hermitian symmetric domains.*

PROOF. Apply Proposition 5.7 and Proposition 4.8. $\square$

Let (G, X) be a Shimura datum. For every $h\colon \mathbb{S} \to G(\mathbb{R})$ in X, $\mathbb{S}$ acts on $\mathrm{Lie}(G)_{\mathbb{C}}$ through the characters $z/\overline{z}$, 1, $\overline{z}/z$. Thus, for $r \in \mathbb{R}^\times \subset \mathbb{C}^\times$, $h(r)$ acts trivially on $\mathrm{Lie}(G)_{\mathbb{C}}$. As the adjoint action of G on $\mathrm{Lie}(G)$ factors through G^{ad} and $\mathrm{Ad}\colon G^{\mathrm{ad}} \to \mathrm{GL}(\mathrm{Lie}(G))$ is injective, this implies that $h(r) \in Z(\mathbb{R})$ where Z is the centre of G. Thus, $h|\mathbb{G}_m$ is independent of h — we denote its reciprocal by w_X (or simply w) and we call w_X the ***weight homomorphism***. For any representation $\rho\colon G_{\mathbb{R}} \to \mathrm{GL}(V)$, $\rho \circ w_X$ defines a decomposition of $V = \bigoplus V_n$ which is the weight decomposition of the hodge structure $(V, \rho \circ h)$ for every $h \in X$.

PROPOSITION 5.9. *Let (G, X) be a Shimura datum. Then X has a unique structure of a complex manifold such that, for every representation $\rho\colon G_{\mathbb{R}} \to \mathrm{GL}(V)$,*

$(V, \rho \circ h)_{h \in X}$ is a holomorphic family of hodge structures. For this complex structure, each family $(V, \rho \circ h)_{h \in X}$ is a variation of hodge structures, and so X is a finite disjoint union of hermitian symmetric domains.

PROOF. Let $\rho\colon G_{\mathbb{R}} \to \mathrm{GL}(V)$ be a faithful representation of $G_{\mathbb{R}}$. The family of hodge structures $(V, \rho \circ h)_{h \in X}$ is continuous, and a slight generalization of (a) of Theorem 2.14 shows that X has a unique structure of a complex manifold for which this family is holomorphic. It follows from Waterhouse 1979, 3.5, that the family of hodge structures defined by every representation is then holomorphic for this complex structure. The condition SV1 implies that $(V, \rho \circ h)_h$ is a variation of hodge structures, and so we can apply (b) of Theorem 2.14. □

Of course, the complex structures defined on X by (5.8) and (5.9) coincide.

ASIDE 5.10. Let (G, X) be a Shimura datum. The maps $\pi_0(X) \to \pi_0(\overline{X})$ and $G(\mathbb{R})/G(\mathbb{R})_+ \to G^{\mathrm{ad}}(\mathbb{R})/G^{\mathrm{ad}}(\mathbb{R})^+$ are injective, and the second can be identified with the first once an $h \in X$ has been chosen. In general, the maps will not be surjective unless $H^1(\mathbb{R}, Z) = 0$.

Shimura varieties. Let (G, X) be a Shimura datum.

LEMMA 5.11. *For any connected component X^+ of X, the natural map*

$$G(\mathbb{Q})_+ \backslash X^+ \times G(\mathbb{A}_f) \to G(\mathbb{Q}) \backslash X \times G(\mathbb{A}_f)$$

is a bijection.

PROOF. Because $G(\mathbb{Q})$ is dense in $G(\mathbb{R})$ (see 5.4) and $G(\mathbb{R})$ acts transitively on X, every $x \in X$ is of the form qx^+ with $q \in G(\mathbb{Q})$ and $x^+ \in X^+$. This shows that the map is surjective.

Let (x, a) and (x', a') be elements of $X^+ \times G(\mathbb{A}_f)$. If $[x, a] = [x', a']$ in $G(\mathbb{Q}) \backslash X \times G(\mathbb{A}_f)$, then

$$x' = qx, \quad a' = qa, \quad \text{some } q \in G(\mathbb{Q}).$$

Because x and x' are both in X^+, q stabilizes X^+ and so lies in $G(\mathbb{R})_+$ (see 5.7). Therefore, $[x, a] = [x', a']$ in $G(\mathbb{Q})_+ \backslash X \times G(\mathbb{A}_f)$. □

LEMMA 5.12. *For any open subgroup K of $G(\mathbb{A}_f)$, the set $G(\mathbb{Q})_+ \backslash G(\mathbb{A}_f)/K$ is finite.*

PROOF. Since $G(\mathbb{Q})_+ \backslash G(\mathbb{Q}) \to G^{\mathrm{ad}}(\mathbb{R})^+ \backslash G^{\mathrm{ad}}(\mathbb{R})$ is injective and the second group is finite (5.3), it suffices to show that $G(\mathbb{Q}) \backslash G(\mathbb{A}_f)/K$ is finite. Later (Theorem 5.17) we shall show that this follows from the strong approximation theorem if G^{der} is simply connected, and the general case is not much more difficult. □

For K a compact open subgroup K of $G(\mathbb{A}_f)$, consider the double coset space

$$\mathrm{Sh}_K(G, X) = G(\mathbb{Q}) \backslash X \times G(\mathbb{A}_f)/K$$

in which $G(\mathbb{Q})$ acts on X and $G(\mathbb{A}_f)$ on the left, and K acts on $G(\mathbb{A}_f)$ on the right:

$$q(x, a)k = (qx, qak), \quad q \in G(\mathbb{Q}), \quad x \in X, \quad a \in G(\mathbb{A}_f), \quad k \in K.$$

LEMMA 5.13. *Let $\mathcal{C}$ be a set of representatives for the double coset space $G(\mathbb{Q})_+ \backslash G(\mathbb{A}_f)/K$, and let X^+ be a connected component of X. Then*

$$G(\mathbb{Q}) \backslash X \times G(\mathbb{A}_f)/K \cong \textstyle\bigsqcup_{g \in \mathcal{C}} \Gamma_g \backslash X^+$$

where Γ_g is the subgroup $gKg^{-1} \cap G(\mathbb{Q})_+$ of $G(\mathbb{Q})_+$. When we endow X with its usual topology and $G(\mathbb{A}_f)$ with its adèlic topology (equivalently, the discrete topology), this becomes a homeomorphism.

PROOF. It is straightforward to prove that, for $g \in \mathcal{C}$, the map

$$[x] \mapsto [x, g] \colon \Gamma_g \backslash X^+ \to G(\mathbb{Q})_+ \backslash X^+ \times G(\mathbb{A}_f)/K$$

is injective, and that $G(\mathbb{Q})_+ \backslash X^+ \times G(\mathbb{A}_f)/K$ is the disjoint union of the images of these maps. Thus, the first statement follows from (5.11). The second statement can be proved in the same way as the similar statement in (4.18). □

Because Γ_g is a congruence subgroup of $G(\mathbb{Q})$, its image in $G^{\mathrm{ad}}(\mathbb{Q})$ is arithmetic (3.2), and so (by definition) its image in $\mathrm{Aut}(X^+)$ is arithmetic. Moreover, when K is sufficiently small, Γ_g will be neat for all $g \in \mathcal{C}$ (apply 3.5) and so its image in $\mathrm{Aut}(X^+)^+$ will also be neat and hence torsion free. Then $\Gamma_g \backslash X^+$ is an arithmetic locally symmetric variety, and $\mathrm{Sh}_K(G, X)$ is finite disjoint of such varieties. Moreover, for an inclusion $K' \subset K$ of sufficiently small compact open subgroups of $G(\mathbb{A}_f)$, the natural map $\mathrm{Sh}_{K'}(G, X) \to \mathrm{Sh}_K(G, X)$ is regular. Thus, when we vary K (sufficiently small), we get an inverse system of algebraic varieties $(\mathrm{Sh}_K(G, X))_K$. There is a natural action of $G(\mathbb{A}_f)$ on the system: for $g \in G(\mathbb{A}_f)$, $K \mapsto g^{-1}Kg$ maps compact open subgroups to compact open subgroups, and

$$\mathcal{T}(g) \colon\ \mathrm{Sh}_K(G, X) \to \mathrm{Sh}_{g^{-1}Kg}(G, X)$$

acts on points as

$$[x, a] \mapsto [x, ag] \colon G(\mathbb{Q}) \backslash X \otimes G(\mathbb{A}_f)/K \to G(\mathbb{Q}) \backslash X \times G(\mathbb{A}_f)/g^{-1}Kg.$$

Note that this is a right action: $\mathcal{T}(gh) = \mathcal{T}(h) \circ \mathcal{T}(g)$.

DEFINITION 5.14. The ***Shimura variety*** $\mathrm{Sh}(G, X)$ ***attached to the Shimura datum*** (G, X) is the inverse system of varieties $(\mathrm{Sh}_K(G, X))_K$ endowed with the action of $G(\mathbb{A}_f)$ described above. Here K runs through the sufficiently small compact open subgroups of $G(\mathbb{A}_f)$.

Morphisms of Shimura varieties.

DEFINITION 5.15. Let (G, X) and (G', X') be Shimura data.

(a) A ***morphism of Shimura data*** $(G, X) \to (G', X')$ is a homomorphism $G \to G'$ of algebraic groups sending X into X'.
(b) A ***morphism of Shimura varieties*** $\mathrm{Sh}(G, X) \to \mathrm{Sh}(G', X')$ is an inverse system of regular maps of algebraic varieties compatible with the action of $G(\mathbb{A}_f)$.

THEOREM 5.16. *A morphism of Shimura data $(G, X) \to (G', X')$ defines a morphism $\mathrm{Sh}(G, X) \to \mathrm{Sh}(G', X')$ of Shimura varieties, which is a closed immersion if $G \to G'$ is injective.*

PROOF. The first part of the statement is obvious from (3.14), and the second is proved in Theorem 1.15 of Deligne 1971*b*. □

The structure of a Shimura variety. By the structure of $\mathrm{Sh}(G,X)$, I mean the structure of the set of connected components and the structure of each connected component. This is worked out in general in Deligne 1979, 2.1.16, but the result there is complicated. When G^{der} is simply connected,[12] it is possible to prove a more pleasant result: the set of connected components is a "zero-dimensional Shimura variety", and each connected component is a connected Shimura variety.

Let (G,X) be a Shimura datum. As on p303, Z is the centre of G and T the largest commutative quotient of G. There are homomorphisms $Z \hookrightarrow G \xrightarrow{\nu} T$, and we define

$$\begin{aligned} T(\mathbb{R})^{\dagger} &= \mathrm{Im}(Z(\mathbb{R}) \to T(\mathbb{R})), \\ T(\mathbb{Q})^{\dagger} &= T(\mathbb{Q}) \cap T(\mathbb{R})^{\dagger}. \end{aligned}$$

Because $Z \to T$ is surjective, $T(\mathbb{R})^{\dagger} \supset T(\mathbb{R})^{+}$ (see 5.1), and so $T(\mathbb{R})^{\dagger}$ and $T(\mathbb{Q})^{\dagger}$ are of finite index in $T(\mathbb{R})$ and $T(\mathbb{Q})$ (see 5.3). For example, for $G = \mathrm{GL}_2$, $T(\mathbb{Q})^{\dagger} = T(\mathbb{Q})^{+} = \mathbb{Q}_{>0}$.

THEOREM 5.17. *Assume G^{der} is simply connected. For K sufficiently small, the natural map*

$$G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K \to T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K)$$

defines an isomorphism

$$\pi_0(\mathrm{Sh}_K(G,X)) \cong T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K).$$

Moreover, $T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K)$ is finite, and the connected component over [1] *is canonically isomorphic to $\Gamma\backslash X^{+}$ for some congruence subgroup Γ of $G^{der}(\mathbb{Q})$ containing $K \cap G^{der}(\mathbb{Q})$.*

In Lemma 5.20 below, we show that $\nu(G(\mathbb{Q})_{+}) \subset T(\mathbb{Q})^{\dagger}$. The "natural map" in the theorem is

$$G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K \overset{5.11}{\cong} G(\mathbb{Q})_{+}\backslash X^{+} \times G(\mathbb{A}_f)/K \xrightarrow{[x,g]\mapsto[\nu(g)]} T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K).$$

The theorem gives a diagram

$$\begin{tikzcd} G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K \arrow[d] & \Gamma\backslash X^{+}, \arrow[l, hook'] \arrow[d] \\ T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K) & {[1]} \arrow[l, hook'] \end{tikzcd}$$

in which $T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K)$ is finite and discrete, the left hand map is continuous and onto with connected fibres, and $\Gamma\backslash X^{+}$ is the fibre over [1].

LEMMA 5.18. *Assume G^{der} is simply connected. Then $G(\mathbb{R})_{+} = G^{der}(\mathbb{R})\cdot Z(\mathbb{R})$.*

---

[12]The Shimura varieties with simply connected derived group are the most important — if one knows everything about them, then one knows everything about all Shimura varieties (because the remainder are quotients of them). However, there are naturally occurring Shimura varieties for which G^{der} is not simply connected, and so we should not ignore them.

Proof. Because G^{der} is simply connected, $G^{\mathrm{der}}(\mathbb{R})$ is connected (5.2) and so $G^{\mathrm{der}}(\mathbb{R}) \subset G(\mathbb{R})_+$. Hence $G(\mathbb{R})_+ \supset G^{\mathrm{der}}(\mathbb{R}) \cdot Z(\mathbb{R})$. For the converse, we use the exact commutative diagram:

$$\begin{array}{ccccccccc}
1 & \longrightarrow & Z'(\mathbb{R}) & \xrightarrow{z\mapsto(z^{-1},z)} & Z(\mathbb{R})\times G^{\mathrm{der}}(\mathbb{R}) & \xrightarrow{(z,g)\mapsto zg} & G(\mathbb{R}) & \longrightarrow & H^1(\mathbb{R},Z') \\
 & & \| & & \downarrow{\scriptstyle (z,g)\mapsto g} & & \downarrow & & \| \\
1 & \longrightarrow & Z'(\mathbb{R}) & \longrightarrow & G^{\mathrm{der}}(\mathbb{R}) & \longrightarrow & G^{\mathrm{ad}}(\mathbb{R}) & \longrightarrow & H^1(\mathbb{R},Z').
\end{array}$$

As $G^{\mathrm{der}} \to G^{\mathrm{ad}}$ is surjective, so also is $G^{\mathrm{der}}(\mathbb{R}) \to G^{\mathrm{ad}}(\mathbb{R})^+$ (see 5.1). Therefore, an element g of $G(\mathbb{R})$ lies in $G(\mathbb{R})_+$ if and only if its image in $G^{\mathrm{ad}}(\mathbb{R})$ lifts to $G^{\mathrm{der}}(\mathbb{R})$. Thus,

$$\begin{aligned}
g \in G(\mathbb{R})_+ &\iff g \mapsto 0 \text{ in } H^1(\mathbb{R}, Z') \\
&\iff g \text{ lifts to } Z(\mathbb{R}) \times G^{\mathrm{der}}(\mathbb{R}) \\
&\iff g \in Z(\mathbb{R}) \cdot G^{\mathrm{der}}(\mathbb{R})
\end{aligned}$$
□

Lemma 5.19. *Let H be a simply connected semisimple algebraic group H over $\mathbb{Q}$.*

(a) *For every finite prime, the group $H^1(\mathbb{Q}_\ell, H) = 0$.*
(b) *The map $H^1(\mathbb{Q}, H) \to \prod_{l\leq\infty} H^1(\mathbb{Q}_l, H)$ is injective (Hasse principle).*

Proof. (a) See Platonov and Rapinchuk 1994, Theorem 6.4, p284.
(b) See ibid., Theorem 6.6, p286. □

Both statements fail for groups that are not simply connected.

Lemma 5.20. *Assume G^{der} is simply connected, and let $t \in T(\mathbb{Q})$. Then $t \in T(\mathbb{Q})^\dagger$ if and only if t lifts to an element of $G(\mathbb{Q})_+$.*

Proof. Lemma 5.19 implies that the vertical arrow at right in the following diagram is injective:

$$\begin{array}{ccccccccc}
1 & \longrightarrow & G^{\mathrm{der}}(\mathbb{Q}) & \longrightarrow & G(\mathbb{Q}) & \xrightarrow{\nu} & T(\mathbb{Q}) & \longrightarrow & H^1(\mathbb{Q}, G^{\mathrm{der}}) \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow{\scriptstyle \text{injective}} \\
1 & \longrightarrow & G^{\mathrm{der}}(\mathbb{R}) & \longrightarrow & G(\mathbb{R}) & \xrightarrow{\nu} & T(\mathbb{R}) & \longrightarrow & H^1(\mathbb{R}, G^{\mathrm{der}})
\end{array}$$

Let $t \in T(\mathbb{Q})^\dagger$. By definition, the image $t_\mathbb{R}$ of t in $T(\mathbb{R})$ lifts to an element $z \in Z(\mathbb{R}) \subset G(\mathbb{R})$. From the diagram, we see that this implies that t maps to the trivial element in $H^1(\mathbb{Q}, G^{\mathrm{der}})$ and so it lifts to an element $g \in G(\mathbb{Q})$. Now $g_\mathbb{R} \cdot z^{-1} \mapsto t_\mathbb{R} \cdot t_\mathbb{R}^{-1} = 1$ in $T(\mathbb{R})$, and so $g_\mathbb{R} \in G^{\mathrm{der}}(\mathbb{R}) \cdot z \subset G^{\mathrm{der}}(\mathbb{R}) \cdot Z(\mathbb{R}) \subset G(\mathbb{R})_+$. Therefore, $g \in G(\mathbb{Q})_+$.

Let t be an element of $T(\mathbb{Q})$ lifting to an element a of $G(\mathbb{Q})_+$. According to 5.18, $a_\mathbb{R} = gz$ for some $g \in G^{\mathrm{der}}(\mathbb{R})$ and $z \in Z(\mathbb{R})$. Now $a_\mathbb{R}$ and z map to the same element in $T(\mathbb{R})$, namely, to $t_\mathbb{R}$, and so $t \in T(\mathbb{Q})^\dagger$ □

The lemma allows us to write

$$T(\mathbb{Q})^\dagger \backslash T(\mathbb{A}_f)/\nu(K) = \nu(G(\mathbb{Q})_+)\backslash T(\mathbb{A}_f)/\nu(K).$$

We now study the fibre over [1] of the map

$$G(\mathbb{Q})_+\backslash X^+ \times G(\mathbb{A}_f)/K \xrightarrow{[x,g]\mapsto[\nu(g)]} \nu(G(\mathbb{Q})_+)\backslash T(\mathbb{A}_f)/\nu(K).$$

Let $g \in G(\mathbb{A}_f)$. If $[\nu(g)] = [1]_K$, then $\nu(g) = \nu(q)\nu(k)$ some $q \in G(\mathbb{Q})_+$ and $k \in K$. It follows that $\nu(q^{-1}gk^{-1}) = 1$, that $q^{-1}gk^{-1} \in G^{\mathrm{der}}(\mathbb{A}_f)$, and that $g \in G(\mathbb{Q})_+ \cdot G^{\mathrm{der}}(\mathbb{A}_f) \cdot K$. Hence every element of the fibre over [1] is represented by an element (x, a) with $a \in G^{\mathrm{der}}(\mathbb{A}_f)$. But, according to the strong approximation theorem (4.16), $G^{\mathrm{der}}(\mathbb{A}_f) = G^{\mathrm{der}}(\mathbb{Q}) \cdot (K \cap G^{\mathrm{der}}(\mathbb{A}_f))$, and so the fibre over [1] is a quotient of X^+; in particular, it is connected. More precisely, it equals $\Gamma\backslash X^+$ where Γ is the image of $K \cap G(\mathbb{Q})_+$ in $G^{\mathrm{ad}}(\mathbb{Q})^+$. This Γ is an arithmetic subgroup of $G^{\mathrm{ad}}(\mathbb{Q})^+$ containing the image of the congruence subgroup $K \cap G^{\mathrm{der}}(\mathbb{Q})$ of $G^{\mathrm{der}}(\mathbb{Q})$. Moreover, arbitrarily small such Γ's arise in this way. Hence, the inverse system of fibres over [1] (indexed by the compact open subgroups K of $G(\mathbb{A}_f)$) is equivalent to the inverse system $\mathrm{Sh}^\circ(G^{\mathrm{der}}, X^+) = (\Gamma\backslash X^+)$.

The study of the fibre over $[t]$ will be similar once we show that there exists an $a \in G(\mathbb{A}_f)$ mapping to t (so that the fibre is nonempty). This follows from the next lemma.

LEMMA 5.21. *Assume G^{der} is simply connected. Then the map $\nu\colon G(\mathbb{A}_f) \to T(\mathbb{A}_f)$ is surjective and sends compact open subgroups to compact open subgroups.*

PROOF. We have to show:

(a) the homomorphism $\nu\colon G(\mathbb{Q}_\ell) \to T(\mathbb{Q}_\ell)$ is surjective for all finite ℓ;
(b) the homomorphism $\nu\colon G(\mathbb{Z}_\ell) \to T(\mathbb{Z}_\ell)$ is surjective for almost all ℓ.

(a) For each prime ℓ, there is an exact sequence

$$1 \to G^{\mathrm{der}}(\mathbb{Q}_\ell) \to G(\mathbb{Q}_\ell) \xrightarrow{\nu} T(\mathbb{Q}_\ell) \to H^1(\mathbb{Q}_\ell, G^{\mathrm{der}})$$

and so (5.19a) shows that $\nu\colon G(\mathbb{Q}_\ell) \to T(\mathbb{Q}_\ell)$ is surjective.

(b) Extend the homomorphism $G \to T$ to a homomorphism of group schemes $\mathcal{G} \to \mathcal{T}$ over $\mathbb{Z}[\frac{1}{N}]$ for some integer N. After N has been enlarged, this map will be a smooth morphism of group schemes and its kernel $\mathcal{G}'$ will have nonsingular connected fibres. On extending the base ring to $\mathbb{Z}_\ell$, $\ell \nmid N$, we obtain an exact sequence

$$0 \to \mathcal{G}'_\ell \to \mathcal{G}_\ell \xrightarrow{\nu} \mathcal{T}_\ell \to 0$$

of group schemes over $\mathbb{Z}_\ell$ such that ν is smooth and $(\mathcal{G}'_\ell)_{\mathbb{F}_\ell}$ is nonsingular and connected. Let $P \in \mathcal{T}_\ell(\mathbb{Z}_\ell)$, and let $Y = \nu^{-1}(P) \subset \mathcal{G}_\ell$. We have to show that $Y(\mathbb{Z}_\ell)$ is nonempty. By Lang's lemma (Springer 1998, 4.4.17), $H^1(\mathbb{F}_\ell, (\mathcal{G}'_\ell)_{\mathbb{F}_\ell}) = 0$, and so

$$\nu\colon \mathcal{G}_\ell(\mathbb{F}_\ell) \to \mathcal{T}_\ell(\mathbb{F}_\ell)$$

is surjective. Therefore $Y(\mathbb{F}_\ell)$ is nonempty. Because Y is smooth over $\mathbb{Z}_\ell$, an argument as in the proof of Newton's lemma (e.g., ANT 7.22) now shows that a point $Q_0 \in Y(\mathbb{F}_\ell)$ lifts to a point $Q \in Y(\mathbb{Z}_\ell)$. □

It remains to show that $T(\mathbb{Q})^\dagger\backslash T(\mathbb{A}_f)/\nu(K)$ is finite. Because $T(\mathbb{Q})^\dagger$ has finite index in $T(\mathbb{Q})$, it suffices to prove that $T(\mathbb{Q})\backslash T(\mathbb{A}_f)/\nu(K)$ is finite. But $\nu(K)$ is open, and so this follows from the next lemma.

LEMMA 5.22. *For any torus T over $\mathbb{Q}$, $T(\mathbb{Q})\backslash T(\mathbb{A}_f)$ is compact.*

PROOF. Consider first the case $T = \mathbb{G}_m$. Then

$$T(\mathbb{A}_f)/T(\hat{\mathbb{Z}}) = \mathbb{A}_f^\times/\hat{\mathbb{Z}}^\times \cong \bigoplus\nolimits_{\ell \text{ finite}} \mathbb{Q}_\ell^\times/\mathbb{Z}_\ell^\times \xrightarrow[\cong]{\oplus \mathrm{ord}_\ell} \bigoplus\nolimits_{\ell \text{ finite}} \mathbb{Z},$$

which is the group of fractional ideals of $\mathbb{Z}$. Therefore, $\mathbb{Q}^\times\backslash\mathbb{A}_f^\times/\hat{\mathbb{Z}}^\times$ is the ideal class group of $\mathbb{Z}$, which is trivial: $\mathbb{A}_f^\times = \mathbb{Q}^\times\cdot\hat{\mathbb{Z}}^\times$. Hence $\mathbb{Q}^\times\backslash\mathbb{A}_f^\times$ is a quotient of $\hat{\mathbb{Z}}^\times$, which is compact.

For a number field F, the same argument using the finiteness of the class number of F shows that $F^\times\backslash\mathbb{A}_{F,f}^\times$ is compact. Here $\mathbb{A}_{F,f}^\times = \prod_{v\text{ finite}}(F_v^\times : \mathcal{O}_v^\times)$.

An arbitrary torus T over $\mathbb{Q}$ will split over some number field, say, $T_F \approx \mathbb{G}_m^{\dim(T)}$. Then $T(F)\backslash T(\mathbb{A}_{F,f}) \approx (F^\times\backslash\mathbb{A}_{F,f}^\times)^{\dim(T)}$, which is compact, and $T(\mathbb{Q})\backslash T(\mathbb{A}_f)$ is a closed subset of it. □

REMARK 5.23. One may ask whether the fibre over [1] equals

$$\Gamma\backslash X^+ = G^{\mathrm{der}}(\mathbb{Q})\backslash X^+ \times G^{\mathrm{der}}(\mathbb{A}_f)/K\cap G^{\mathrm{der}}(\mathbb{A}_f),\quad \Gamma = K\cap G^{\mathrm{der}}(\mathbb{Q}),$$

rather than quotient of X^+ by some larger group than Γ. This will be true if Z' satisfies the Hasse principle for H^1 (for then every element in $G(\mathbb{Q})_+\cap K$ with K sufficiently small will lie in $G^{\mathrm{der}}(\mathbb{Q})\cdot Z(\mathbb{Q})$). It is known that Z' satisfies the Hasse principle for H^1 when G^{der} has no isogeny factors of type A, but not in general otherwise (Milne 1987). This is one reason why, in the definition of $\mathrm{Sh}^\circ(G^{\mathrm{der}}, X^+)$, we include quotients $\Gamma\backslash X^+$ in which Γ is an arithmetic subgroup of $G^{\mathrm{ad}}(\mathbb{Q})^+$ containing, but not necessarily equal to, the image of congruence subgroup of $G^{\mathrm{der}}(\mathbb{Q})$.

Zero-dimensional Shimura varieties. Let T be a torus over $\mathbb{Q}$. According to Deligne's definition, every homomorphism $h\colon \mathbb{C}^\times \to T(\mathbb{R})$ defines a Shimura variety $\mathrm{Sh}(T,\{h\})$ — in this case the conditions SV1,2,3 are vacuous. For any compact open $K\subset T(\mathbb{A}_f)$,

$$\mathrm{Sh}_K(T,\{h\}) = T(\mathbb{Q})\backslash\{h\}\times T(\mathbb{A}_f)/K \cong T(\mathbb{Q})\backslash T(\mathbb{A}_f)/K$$

(finite discrete set). We should extend this definition a little. Let Y be a finite set on which $T(\mathbb{R})/T(\mathbb{R})^+$ acts transitively. Define $\mathrm{Sh}(T,Y)$ to be the inverse system of finite sets

$$\mathrm{Sh}_K(T,Y) = T(\mathbb{Q})\backslash Y\times T(\mathbb{A}_f)/K,$$

with K running over the compact open subgroups of $T(\mathbb{A}_f)$. Call such a system a ***zero-dimensional Shimura variety.***

Now let (G,X) be a Shimura datum with G^{der} simply connected, and let $T = G/G^{\mathrm{der}}$. Let $Y = T(\mathbb{R})/T(\mathbb{R})^\dagger$. Because $T(\mathbb{Q})$ is dense in $T(\mathbb{R})$ (see 5.4), $Y\cong T(\mathbb{Q})/T(\mathbb{Q})^\dagger$ and

$$T(\mathbb{Q})^\dagger\backslash T(\mathbb{A}_f)/K \cong T(\mathbb{Q})\backslash Y\times T(\mathbb{A}_f)/K$$

Thus, we see that if G^{der} is simply connected, then

$$\pi_0(\mathrm{Sh}_K(G,X)) \cong \mathrm{Sh}_{\nu(K)}(T,Y).$$

In other words, the set of connected components of the Shimura variety is a zero-dimensional Shimura variety (as promised).

Additional axioms. The weight homomorphism w_X is a homomorphism $\mathbb{G}_m\to G_{\mathbb{R}}$ over $\mathbb{R}$ of algebraic groups that are defined over $\mathbb{Q}$. It is therefore defined over $\mathbb{Q}^{\mathrm{al}}$. Some simplifications to the theory occur when some of the following conditions hold:

SV4: The weight homomorphism $w_X\colon \mathbb{G}_m\to G_{\mathbb{R}}$ is defined over $\mathbb{Q}$ (we then say that ***the weight is rational).***

SV5: The group $Z(\mathbb{Q})$ is discrete in $Z(\mathbb{A}_f)$.

SV6: The torus Z° splits over a CM-field (see p334 for the notion of a CM-field).

Let $G \to \mathrm{GL}(V)$ be a representation of G (meaning, of course, a $\mathbb{Q}$-representation). Each $h \in X$ defines a hodge structure on $V(\mathbb{R})$. When SV4 holds, these are rational hodge structures (p283). It is hoped that these hodge structures all occur in the cohomology of algebraic varieties and, moreover, that the Shimura variety is a moduli variety for motives when SV4 holds and a fine moduli variety when additionally SV5 holds. This will be discussed in more detail later. In Theorem 5.26 below, we give a criterion for SV5 to hold.

Axiom SV6 makes some statements more natural. For example, when SV6 holds, w is defined over a totally real field.

EXAMPLE 5.24. Let B be a quaternion algebra over a totally real field F, and let G be the algebraic group over $\mathbb{Q}$ with $G(\mathbb{Q}) = B^{\times}$. Then, $B \otimes_{\mathbb{Q}} F = \prod_v B \otimes_{F,v} \mathbb{R}$ where v runs over the embeddings of F into $\mathbb{R}$. Thus,

$$\begin{array}{ccccccccccc} B \otimes_{\mathbb{Q}} \mathbb{R} & \approx & \mathbb{H} & \times & \cdots & \times & \mathbb{H} & \times & M_2(\mathbb{R}) & \times & \cdots & \times & M_2(\mathbb{R}) \\ G(\mathbb{R}) & \approx & \mathbb{H}^{\times} & \times & \cdots & \times & \mathbb{H}^{\times} & \times & \mathrm{GL}_2(\mathbb{R}) & \times & \cdots & \times & \mathrm{GL}_2(\mathbb{R}) \\ h(a+ib) & = & 1 & & \cdots & & 1 & & \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) & & \cdots & & \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) \\ w(r) & = & 1 & & \cdots & & 1 & & r^{-1}I_2 & & \cdots & & r^{-1}I_2 \end{array}$$

Let X be the $G(\mathbb{R})$-conjugacy class of h. Then (G, X) satisfies SV1 and SV2, and so it is a Shimura datum if B splits at at least one real prime of F. Let $I = \mathrm{Hom}(F, \mathbb{Q}^{\mathrm{al}}) = \mathrm{Hom}(F, \mathbb{R})$, and let I_{nc} be the set of v such that $B \otimes_{F,v} \mathbb{R}$ is split. Then w is defined over the subfield of $\mathbb{Q}^{\mathrm{al}}$ fixed by the automorphisms of $\mathbb{Q}^{\mathrm{al}}$ stabilizing I_{nc}. This field is always totally real, and it equals $\mathbb{Q}$ if and only if $I = I_{\mathrm{nc}}$.

Arithmetic subgroups of tori. Let T be a torus over $\mathbb{Q}$, and let $T(\mathbb{Z})$ be an arithmetic subgroup of $T(\mathbb{Q})$, for example,

$$T(\mathbb{Z}) = \mathrm{Hom}(X^*(T), \mathcal{O}_L^{\times})^{\mathrm{Gal}(L/\mathbb{Q})},$$

where L is some galois splitting field of T. The congruence subgroup problem is known to have a positive answer for tori (Serre 1964, 3.5), i.e., every subgroup of $T(\mathbb{Z})$ of finite index contains a congruence subgroup. Thus the topology induced on $T(\mathbb{Q})$ by that on $T(\mathbb{A}_f)$ has the following description: $T(\mathbb{Z})$ is open, and the induced topology on $T(\mathbb{Z})$ is the profinite topology. In particular,

$$T(\mathbb{Q}) \text{ is discrete } \iff T(\mathbb{Z}) \text{ is discrete } \iff T(\mathbb{Z}) \text{ is finite.}$$

EXAMPLE 5.25. (a) Let $T = \mathbb{G}_m$. Then $T(\mathbb{Z}) = \{\pm 1\}$, and so $T(\mathbb{Q})$ is discrete in $T(\mathbb{A}_f)$. This, of course, can be proved directly.

(b) Let $T(\mathbb{Q}) = \{a \in \mathbb{Q}[\sqrt{-1}]^{\times} \mid \mathrm{Nm}(a) = 1\}$. Then $T(\mathbb{Z}) = \{\pm 1, \pm\sqrt{-1}\}$, and so $T(\mathbb{Q})$ is discrete.

(c) Let $T(\mathbb{Q}) = \{a \in \mathbb{Q}[\sqrt{2}]^{\times} \mid \mathrm{Nm}(a) = 1\}$. Then $T(\mathbb{Z}) = \{\pm(1+\sqrt{2})^n \mid n \in \mathbb{Z}\}$, and so neither $T(\mathbb{Z})$ nor $T(\mathbb{Q})$ is discrete.

THEOREM 5.26. *Let T be a torus over $\mathbb{Q}$, and let $T^a = \bigcap_{\chi} \mathrm{Ker}(\chi\colon T \to \mathbb{G}_m)$ (characters χ of T rational over $\mathbb{Q}$). Then $T(\mathbb{Q})$ is discrete in $T(\mathbb{A}_f)$ if and only if $T^a(\mathbb{R})$ is compact.*

Proof. According to a theorem of Ono (Serre 1968, pII-39), $T(\mathbb{Z}) \cap T^a(\mathbb{Q})$ is of finite index in $T(\mathbb{Z})$, and the quotient $T^a(\mathbb{R})/T(\mathbb{Z}) \cap T^a(\mathbb{Q})$ is compact. Now $T(\mathbb{Z}) \cap T^a(\mathbb{Q})$ is an arithmetic subgroup of $T^a(\mathbb{Q})$, and hence is discrete in $T^a(\mathbb{R})$. It follows that $T(\mathbb{Z}) \cap T^a(\mathbb{Q})$ is finite if and only if $T^a(\mathbb{R})$ is compact. □

For example, in (5.25)(a), $T^a = 1$ and so certainly $T^a(\mathbb{R})$ is compact; in (b), $T^a(\mathbb{R}) = U_1$, which is compact; in (c), $T^a = T$ and $T(\mathbb{R}) = \{(a,b) \in \mathbb{R}\times\mathbb{R} \mid ab = 1\}$, which is not compact.

Remark 5.27. A torus T over a field k is said to be ***anisotropic*** if there are no characters $\chi\colon T \to \mathbb{G}_m$ defined over k. A real torus is anisotropic if and only if it is compact. The torus $T^a =_{\text{df}} \bigcap \operatorname{Ker}(\chi\colon T \to \mathbb{G}_m)$ is the largest anisotropic subtorus of T. Thus (5.26) says that $T(\mathbb{Q})$ is discrete in $T(\mathbb{A}_f)$ if and only if the largest anisotropic subtorus of T remains anisotropic over $\mathbb{R}$.

Note that SV5 holds if and only if $(Z^{\circ a})_{\mathbb{R}}$ is anisotropic.

Let T be a torus that splits over CM-field L. In this case there is a torus $T^+ \subset T$ such that $T^+_L = \bigcap_{\iota\chi=-\chi} \operatorname{Ker}(\chi\colon T_L \to \mathbb{G}_m)$. Then $T(\mathbb{Q})$ is discrete in $T(\mathbb{A}_f)$ if and only if T^+ is split, i.e., if and only if the largest subtorus of T that splits over $\mathbb{R}$ is already split over $\mathbb{Q}$.

Passage to the limit. Let K be a compact open subgroup of $G(\mathbb{A}_f)$, and let $Z(\mathbb{Q})^-$ be the closure of $Z(\mathbb{Q})$ in $Z(\mathbb{A}_f)$. Then $Z(\mathbb{Q}) \cdot K = Z(\mathbb{Q})^- \cdot K$ (in $G(\mathbb{A}_f)$) and

$$\begin{aligned}\operatorname{Sh}_K(G,X) &=_{\text{df}} G(\mathbb{Q})\backslash X \times (G(\mathbb{A}_f)/K)\\ &\cong \frac{G(\mathbb{Q})}{Z(\mathbb{Q})}\Big\backslash X \times (G(\mathbb{A}_f)/Z(\mathbb{Q})\cdot K)\\ &\cong \frac{G(\mathbb{Q})}{Z(\mathbb{Q})}\Big\backslash X \times (G(\mathbb{A}_f)/Z(\mathbb{Q})^-\cdot K).\end{aligned}$$

Theorem 5.28. *For any Shimura datum (G,X),*

$$\varprojlim_K \operatorname{Sh}_K(G,X) = \frac{G(\mathbb{Q})}{Z(\mathbb{Q})}\Big\backslash X \times (G(\mathbb{A}_f)/Z(\mathbb{Q})^-).$$

When SV5 holds,

$$\varprojlim_K \operatorname{Sh}_K(G,X) = G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f).$$

Proof. The first equality can be proved by the same argument as (4.19), and the second follows from the first (cf. Deligne 1979, 2.1.10, 2.1.11). □

Remark 5.29. Put $S_K = \operatorname{Sh}_K(G,X)$. For varying K, the S_K form a variety (scheme) with a right action of $G(\mathbb{A}_f)$ in the sense of Deligne 1979, 2.7.1. This means the following:

(a) the S_K form an inverse system of algebraic varieties indexed by the compact open subgroups K of $G(\mathbb{A}_f)$ (if $K \subset K'$, there is an obvious quotient map $S_{K'} \to S_K$);
(b) there is an action ρ of $G(\mathbb{A}_f)$ on the system $(S_K)_K$ defined by isomorphisms (of algebraic varieties) $\rho_K(a)\colon S_K \to S_{g^{-1}Kg}$ (on points, $\rho_K(a)$ is $[x,a'] \mapsto [x,a'a]$);

(c) for $k \in K$, $\rho_K(k)$ is the identity map; therefore, for K' normal in K, there is an action of the finite group K/K' on $S_{K'}$; the variety S_K is the quotient of $S_{K'}$ by the action of K/K'.

REMARK 5.30. When we regard the $\mathrm{Sh}_K(G,X)$ as schemes, the inverse limit of the system $\mathrm{Sh}_K(G,X)$ exists:

$$S = \varprojlim \mathrm{Sh}_K(G,X).$$

This is a scheme over $\mathbb{C}$, not(!) of finite type, but it is locally noetherian and regular (cf. Milne 1992, 2.4). There is a right action of $G(\mathbb{A}_f)$ on S, and, for K a compact open subgroup of $G(\mathbb{A}_f)$,

$$\mathrm{Sh}_K(G,X) = S/K$$

(Deligne 1979, 2.7.1). Thus, the system $(\mathrm{Sh}_K(G,X))_K$ together with its right action of $G(\mathbb{A}_f)$ can be recovered from S with its right action of $G(\mathbb{A}_f)$. Moreover,

$$S(\mathbb{C}) \cong \varprojlim \mathrm{Sh}_K(G,X)(\mathbb{C}) = \varprojlim G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K.$$

NOTES. Axioms SV1, SV2, SV3, and SV4 are respectively the conditions (2.1.1.1), (2.1.1.2), (2.1.1.3), and (2.1.1.4) of Deligne 1979. Axiom SV5 is weaker than the condition (2.1.1.5) ibid., which requires that $\mathrm{ad}h(i)$ be a Cartan involution on $(G/w(\mathbb{G}_m))_\mathbb{R}$, i.e., that $(Z^\circ/w(\mathbb{G}_m))_\mathbb{R}$ be anisotropic.

6. The Siegel modular variety

In this section, we study the most important Shimura variety, namely, the Siegel modular variety.

Dictionary. Let V be an $\mathbb{R}$-vector space. Recall (2.4) that to give a $\mathbb{C}$-structure J on V is the same as to give a hodge structure h_J on V of type $(-1,0),(0,-1)$. Here h_J is the restriction to $\mathbb{C}^\times$ of the homomorphism

$$a+bi \mapsto a+bJ\colon \mathbb{C} \to \mathrm{End}_\mathbb{R}(V).$$

For the hodge decompostion $V(\mathbb{C}) = V^{-1,0} \oplus V^{-1,0}$,

	$V^{-1,0}$	$V^{0,-1}$
J acts as	$+i$	$-i$
$h_J(z)$ acts as	z	$\overline{z}$

Let ψ be a nondegenerate $\mathbb{R}$-bilinear alternating form on V. A direct calculation shows that

$$\psi(Ju,Jv) = \psi(u,v) \iff \psi(zu,zv) = |z|^2\psi(u,v) \text{ for all } z \in \mathbb{C}.$$

Let $\psi_J(u,v) = \psi(u,Jv)$. Then

$$\psi(Ju,Jv) = \psi(u,v) \iff \psi_J \text{ is symmetric}$$

and

$$\begin{array}{lcl} \psi(Ju,Jv) = \psi(u,v) \text{ and} & \overset{(2.12)}{\iff} & \psi \text{ is a polarization of the} \\ \psi_J \text{ is positive definite} & & \text{hodge structure } (V,h_J). \end{array}$$

Symplectic spaces. Let k be a field of characteristic $\neq 2$, and let (V, ψ) be a ***symplectic space*** of dimension $2n$ over k, i.e., V is a k-vector space of dimension $2n$ and ψ is a nondegenerate alternating form ψ. A subspace W of V is ***totally isotropic*** if $\psi(W, W) = 0$. A ***symplectic basis*** of V is a basis $(e_{\pm i})_{1 \leq i \leq n}$ such that

$$\psi(e_i, e_{-i}) = 1 \text{ for } 1 \leq i \leq n,$$
$$\psi(e_i, e_j) = 0 \text{ for } \quad j \neq \pm i.$$

LEMMA 6.1. *Let W be a totally isotropic subspace of V. Then any basis of W can be extended to a symplectic basis for V. In particular, V has symplectic bases (and two symplectic spaces of the same dimension are isomorphic).*

PROOF. Standard. □

Thus, a maximal totally isotropic subspace of V will have dimension n. Such subspaces are called ***lagrangians***.

Let $\mathrm{GSp}(\psi)$ be the group of ***symplectic similitudes*** of (V, ψ), i.e., the group of automorphisms of V preserving ψ up to a scalar. Thus

$$\mathrm{GSp}(\psi)(k) = \{g \in \mathrm{GL}(V) \mid \psi(gu, gv) = \nu(g) \cdot \psi(u, v) \text{ some } \nu(g) \in k^\times\}.$$

Define $\mathrm{Sp}(\psi)$ by the exact sequence

$$1 \to \mathrm{Sp}(\psi) \to \mathrm{GSp}(\psi) \xrightarrow{\nu} \mathbb{G}_m \to 1.$$

Then $\mathrm{GSp}(\psi)$ has derived group $\mathrm{Sp}(\psi)$, centre $\mathbb{G}_m$, and adjoint group $\mathrm{GSp}(\psi)/\mathbb{G}_m = \mathrm{Sp}(\psi)/\pm I$.

For example, when V has dimension 2, there is only one nondegenerate alternating form on V up to scalars, which must therefore be preserved up to scalars by any automorphism, and so $\mathrm{GSp}(\psi) = \mathrm{GL}_2$ and $\mathrm{Sp}(\psi) = \mathrm{SL}_2$.

The group $\mathrm{Sp}(\psi)$ acts simply transitively on the set of symplectic bases: if $(e_{\pm i})$ and $(f_{\pm i})$ are bases of V, then there is a unique $g \in \mathrm{GL}_{2n}(k)$ such that $ge_{\pm i} = f_{\pm i}$, and if $(e_{\pm i})$ and $(f_{\pm i})$ are both symplectic, then $g \in \mathrm{Sp}(\psi)$.

The Shimura datum attached to a symplectic space. Fix a symplectic space (V, ψ) over $\mathbb{Q}$, and let $G = \mathrm{GSp}(\psi)$ and $S = \mathrm{Sp}(\psi) = G^{\mathrm{der}}$.

Let J be a complex structure on $V(\mathbb{R})$ such that $\psi(Ju, Jv) = \psi(u, v)$. Then $J \in S(\mathbb{R})$, and $h_J(z)$ lies in $G(\mathbb{R})$ (and in $S(\mathbb{R})$ if $|z| = 1$) — see the dictionary. We say that J is ***positive*** (resp. ***negative***) if $\psi_J(u, v) =_{\mathrm{df}} \psi(u, Jv)$ is positive definite (resp. negative definite).

Let X^+ (resp. X^-) denote the set of positive (resp. negative) complex structures on $V(\mathbb{R})$, and let $X = X^+ \sqcup X^-$. Then $G(\mathbb{R})$ acts on X according to the rule

$$(g, J) \mapsto gJg^{-1},$$

and the stabilizer in $G(\mathbb{R})$ of X^+ is

$$G(\mathbb{R})^+ = \{g \in G(\mathbb{R}) \mid \nu(g) > 0\}.$$

For a symplectic basis $(e_{\pm i})$ of V, define J by $Je_{\pm i} = \pm e_{\mp i}$, i.e.,

$$e_i \overset{J}{\longmapsto} e_{-i} \overset{J}{\longmapsto} -e_i, \quad 1 \leq i \leq n.$$

Then $J^2 = -1$ and $J \in X^+$ — in fact, $(e_i)_i$ is an orthonormal basis for ψ_J. Conversely, if $J \in X^+$, then J has this description relative to any orthonormal basis for the positive definite form ψ_J. The map from symplectic bases to X^+ is equivariant for the actions of $S(\mathbb{R})$. Therefore, $S(\mathbb{R})$ acts transitively on X^+, and $G(\mathbb{R})$ acts transitively on X.

For $J \in X$, let h_J be the corresponding homomorphism $\mathbb{C}^\times \to G(\mathbb{R})$. Then $h_{gJg^{-1}}(z) = gh_J(z)g^{-1}$. Thus $J \mapsto h_J$ identifies X with a $G(\mathbb{R})$-conjugacy class of homomorphisms $h\colon \mathbb{C}^\times \to G(\mathbb{R})$. We check that (G, X) satisfies the axioms SV1–SV6.

(SV1). For $h \in X$, let $V^+ = V^{-1,0}$ and $V^- = V^{0,-1}$, so that $V(\mathbb{C}) = V^+ \oplus V^-$ with $h(z)$ acting on V^+ and V^- as multiplication by z and $\overline{z}$ respectively. Then

$$\begin{array}{ccccccccc} \mathrm{Hom}(V(\mathbb{C}), V(\mathbb{C})) & = & \mathrm{Hom}(V^+, V^+) & \oplus & \mathrm{Hom}(V^+, V^-) & \oplus & \mathrm{Hom}(V^-, V^+) & \oplus & \mathrm{Hom}(V^-, V^-) \\ h(z) \text{ acts as} & & 1 & & z/\overline{z} & & \overline{z}/z & & 1 \end{array}$$

The Lie algebra of G is the subspace

$$\mathrm{Lie}(G) = \{f \in \mathrm{Hom}(V, V) \mid \psi(f(u), v) + \psi(u, f(v)) = 0\},$$

of $\mathrm{End}(V)$, and so SV1 holds.

(SV2). We have to show that $\mathrm{ad}J$ is a Cartan involution on G^{ad}. But, $J^2 = -1$ lies in the centre of $S(\mathbb{R})$ and ψ is a J-polarization for $S_{\mathbb{R}}$ in the sense of (1.20), which shows that $\mathrm{ad}J$ is a Cartan involution for S.

(SV3). In fact, G^{ad} is $\mathbb{Q}$-simple, and $G^{\mathrm{ad}}(\mathbb{R})$ is not compact.

(SV4). For $r \in \mathbb{R}^\times$, $w_h(r)$ acts on both $V^{-1,0}$ and $V^{0,-1}$ as $v \mapsto rv$. Therefore, w_X is the homomorphism $\mathbb{G}_{m\mathbb{R}} \to \mathrm{GL}(V(\mathbb{R}))$ sending $r \in \mathbb{R}^\times$ to multplication by r. This is defined over $\mathbb{Q}$.

(SV5). The centre of G is $\mathbb{G}_m$, and $\mathbb{Q}^\times$ is discrete in $\mathbb{A}_f^\times$ (see 5.25).

(SV6). The centre of G is split already over $\mathbb{Q}$.

We often write $(G(\psi), X(\psi))$ for the Shimura datum defined by a symplectic space (V, ψ), and $(S(\psi), X(\psi)^+)$ for the connected Shimura datum.

EXERCISE 6.2. (a) Show that for any $h \in X(\psi)$, $\nu(h(z)) = z\overline{z}$. [Hint: for nonzero $v^+ \in V^+$ and $v^- \in V^-$, compute $\psi_{\mathbb{C}}(h(z)v^+, h(z)v^-)$ in two different ways.]

(b) Show that the choice of a symplectic basis for V identifies X^+ with $\mathcal{H}_g$ as an $\mathrm{Sp}(\psi)$-set (see 1.2).

The Siegel modular variety. Let $(G, X) = (G(\psi), X(\psi))$ be the Shimura datum defined by a symplectic space (V, ψ) over $\mathbb{Q}$. The ***Siegel modular variety attached to*** (V, ψ) is the Shimura variety $\mathrm{Sh}(G, X)$.

Let $V(\mathbb{A}_f) = \mathbb{A}_f \otimes_{\mathbb{Q}} V$. Then $G(\mathbb{A}_f)$ is the group of $\mathbb{A}_f$-linear automorphisms of $V(\mathbb{A}_f)$ preserving ψ up to multiplication by an element of $\mathbb{A}_f^\times$.

Let K be a compact open subgroup of $G(\mathbb{A}_f)$, and let $\mathcal{H}_K$ be the set of triples $((W, h), s, \eta K)$ where

- (W, h) is a rational hodge structure of type $(-1, 0), (0, -1)$;
- $\pm s$ is a polarization for (W, h);
- ηK is a K-orbit of $\mathbb{A}_f$-linear isomorphisms $V(\mathbb{A}_f) \to W(\mathbb{A}_f)$ sending ψ to an $\mathbb{A}_f^\times$-multiple of s.

An ***isomorphism***

$$((W, h), s, \eta K) \to ((W', h'), s', \eta' K)$$

of triples is an isomorphism $b\colon (W, h) \to (W', h')$ of rational hodge structures such that $b(s) = cs'$ some $c \in \mathbb{Q}^\times$ and $b \circ \eta = \eta' \mod K$.

Note that to give an element of $\mathcal{H}_K$ amounts to giving a symplectic space (W, s) over $\mathbb{Q}$, a complex structure on W that is positive or negative for s, and ηK. The existence of η implies that $\dim W = \dim V$, and so (W, s) and (V, ψ) are isomorphic. Choose an isomorphism $a\colon W \to V$ sending ψ to a $\mathbb{Q}^\times$-multiple of s. Then

$$ah =_{\mathrm{df}} (z \mapsto a \circ h(z) \circ a^{-1})$$

lies in X, and

$$V(\mathbb{A}_f) \xrightarrow{\eta} W(\mathbb{A}_f) \xrightarrow{a} V(\mathbb{A}_f)$$

lies in $G(\mathbb{A}_f)$. Any other isomorphism $a'\colon W \to V$ sending ψ to a multiple of s differs from a by an element of $G(\mathbb{Q})$, say, $a' = q \circ a$ with $q \in G(\mathbb{Q})$. Replacing a with a' only replaces $(ah, a \circ \eta)$ with $(qah, qa \circ \eta)$. Similarly, replacing η with ηk replaces $(ah, a \circ \eta)$ with $(ah, a \circ \eta k)$. Therefore, the map

$$(W \ldots) \mapsto [ah, a \circ \eta]_K \colon \mathcal{H}_K \to G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K$$

is well-defined.

Proposition 6.3. *The set* $\mathrm{Sh}_K(G, X)$ *classifies the triples in* $\mathcal{H}_K$ *modulo isomorphism. More precisely, the map* $(W, \ldots) \mapsto [ah, a \circ \eta]_K$ *defines a bijection*

$$\mathcal{H}_K/{\approx} \to G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K.$$

Proof. It is straightforward to check that the map sends isomorphic triples to the same class, and that two triples are isomorphic if they map to the same class. The map is onto because $[h, g]$ is the image of $((V, h), \psi, gK)$. □

Complex abelian varieties. An ***abelian variety*** A over a field k is a connected projective algebraic variety over k together with a group structure given by regular maps. A one-dimensional abelian variety is an elliptic curve. Happily, a theorem, whose origins go back to Riemann, reduces the study of abelian varieties over $\mathbb{C}$ to multilinear algebra.

Recall that a lattice in a real or complex vector space V is the $\mathbb{Z}$-module generated by an $\mathbb{R}$-basis for V. For a lattice Λ in $\mathbb{C}^n$, make $\mathbb{C}^n/\Lambda$ into a complex manifold by endowing it with the quotient structure. A ***complex torus*** is a complex manifold isomorphic to $\mathbb{C}^n/\Lambda$ for some lattice Λ in $\mathbb{C}^n$.

Note that $\mathbb{C}^n$ is the universal covering space of $M = \mathbb{C}^n/\Lambda$ with Λ as its group of covering transformations, and $\pi_1(M, 0) = \Lambda$ (Hatcher 2002, 1.40). Therefore, (ib. 2A.1)

$$H_1(M, \mathbb{Z}) \cong \Lambda \tag{31}$$

and (Greenberg 1967, 23.14)

$$H^1(M, \mathbb{Z}) \cong \mathrm{Hom}(\Lambda, \mathbb{Z}). \tag{32}$$

Proposition 6.4. *Let* $M = \mathbb{C}^n/\Lambda$. *There is a canonical isomorphism*

$$H^n(M, \mathbb{Z}) \cong \mathrm{Hom}(\textstyle\bigwedge^n \Lambda, \mathbb{Z}),$$

i.e., $H^n(M, \mathbb{Z})$ *is canonically isomorphic to the set of* n*-alternating forms* $\Lambda \times \cdots \times \Lambda \to \mathbb{Z}$.

Proof. From (32), we see that

$$\textstyle\bigwedge^n H^1(M, \mathbb{Z}) \cong \bigwedge^n \mathrm{Hom}(\Lambda, \mathbb{Z}).$$

Since[13]

$$\bigwedge^n \operatorname{Hom}(\Lambda,\mathbb{Z}) \cong \operatorname{Hom}(\bigwedge^n \Lambda,\mathbb{Z}),$$

we see that it suffices to show that cup-product defines an isomorphism

$$\bigwedge^n H^1(M,\mathbb{Z}) \to H^n(M,Z). \tag{33}$$

Let $\mathcal{T}$ be the class of topological manifolds M whose cohomology groups are free $\mathbb{Z}$-modules of finite rank and for which the maps (33) are isomorphisms for all n. Certainly, the circle S^1 is in $\mathcal{T}$ (its cohomology groups are $\mathbb{Z}$, $\mathbb{Z}$, $0,\ldots$), and the Künneth formula (Hatcher 2002, 3.16 et seq.) shows that if M_1 and M_2 are in $\mathcal{T}$, then so also is $M_1 \times M_2$. As a topological manifold, $\mathbb{C}^n/\Lambda \approx (S^1)^{2n}$, and so M is in $\mathcal{T}$. □

PROPOSITION 6.5. *A linear map $\alpha\colon \mathbb{C}^n \to \mathbb{C}^{n'}$ such that $\alpha(\Lambda) \subset \Lambda'$ defines a holomorphic map $\mathbb{C}^n/\Lambda \to \mathbb{C}^{n'}/\Lambda'$ sending 0 to 0, and every holomorphic map $\mathbb{C}^n/\Lambda \to \mathbb{C}^{n'}/\Lambda'$ sending 0 to 0 is of this form (for a unique α).*

PROOF. The map $\mathbb{C}^n \xrightarrow{\alpha} \mathbb{C}^{n'} \to \mathbb{C}^{n'}/\Lambda'$ is holomorphic, and it factors through $\mathbb{C}^n/\Lambda$. Because $\mathbb{C}/\Lambda$ has the quotient structure, the resulting map $\mathbb{C}^n/\Lambda \to \mathbb{C}^{n'}/\Lambda'$ is holomorphic. Conversely, let $\varphi\colon \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$ be a holomorphic map such that $\varphi(0) = 0$. Then $\mathbb{C}^n$ and $\mathbb{C}^{n'}$ are universal covering spaces of $\mathbb{C}^n/\Lambda$ and $\mathbb{C}^{n'}/\Lambda'$, and a standard result in topology (Hatcher 2002, 1.33, 1.34) shows that φ lifts uniquely to a continuous map $\tilde{\varphi}\colon \mathbb{C}^n \to \mathbb{C}^{n'}$ such that $\tilde{\varphi}(0) = 0$:

$$\begin{tikzcd} \mathbb{C}^n \arrow[r, "\tilde{\varphi}"] \arrow[d] & \mathbb{C}^{n'} \arrow[d] \\ \mathbb{C}^n/\Lambda \arrow[r, "\varphi"] & \mathbb{C}^{n'}/\Lambda'. \end{tikzcd}$$

Because the vertical arrows are local isomorphisms, $\tilde{\varphi}$ is automatically holomorphic. For any $\omega \in \Lambda$, the map $z \mapsto \tilde{\varphi}(z+\omega) - \tilde{\varphi}(z)$ is continuous and takes values in $\Lambda' \subset \mathbb{C}$. Because $\mathbb{C}^n$ is connected and Λ' is discrete, it must be constant. Therefore, for each j, $\frac{\partial\tilde{\varphi}}{\partial z_j}$ is a doubly periodic function, and so defines a holomorphic function $\mathbb{C}^n/\Lambda \to \mathbb{C}^{n'}$, which must be constant (because $\mathbb{C}^n/\Lambda$ is compact). Write $\tilde{\varphi}$ as an n'-tuple $(\tilde{\varphi}_1,\ldots,\tilde{\varphi}_{n'})$ of holomorphic functions $\tilde{\varphi}_i$ in n variables. Because $\tilde{\varphi}_i(0) = 0$ and $\frac{\partial\tilde{\varphi}_i}{\partial z_j}$ is constant for each j, the power series expansion of $\tilde{\varphi}_i$ at 0 is of the form $\sum a_{ij} z_j$. Now $\tilde{\varphi}_i$ and $\sum a_{ij} z_j$ are holomorphic functions on $\mathbb{C}^n$ that coincide on a neighbourhood of 0, and so are equal on the whole of $\mathbb{C}^n$. We have shown that

$$\tilde{\varphi}(z_1,\ldots,z_n) = (\textstyle\sum a_{1j}z_j,\ldots,\sum a_{n'j}z_j). \qquad \square$$

ASIDE 6.6. The proposition shows that every holomorphic map $\varphi\colon \mathbb{C}^n/\Lambda \to \mathbb{C}^{n'}/\Lambda'$ such that $\varphi(0) = 0$ is a homomorphism. A similar statement is true for abelian varieties over any field k: a regular map $\varphi\colon A \to B$ of abelian varieties such that $\varphi(0) = 0$ is a homomorphism (AG, 7.14). For example, the map sending an element to its inverse is a homomorphism, which implies that the group law on A

[13]For a free $\mathbb{Z}$-module Λ of finite rank, the pairing

$$\bigwedge^n \Lambda^\vee \times \bigwedge^n \Lambda \to \mathbb{Z}$$

determined by

$$(f_1 \wedge \cdots \wedge f_n, v_1 \otimes \cdots \otimes v_n) = \det(f_i(v_j))$$

is nondegenerate (since it is modulo p for every p — see Bourbaki 1958, §8).

is commutative. Also, the group law on an abelian variety is uniquely determined by the zero element.

Let $M=\mathbb{C}^n/\Lambda$ be a complex torus. The isomorphism $\mathbb{R}\otimes\Lambda\cong\mathbb{C}^n$ defines a complex structure J on $\mathbb{R}\otimes\Lambda$. A ***riemann form*** for M is an alternating form $\psi\colon\Lambda\times\Lambda\to\mathbb{Z}$ such that $\psi_{\mathbb{R}}(Ju,Jv)=\psi_{\mathbb{R}}(u,v)$ and $\psi_{\mathbb{R}}(u,Ju)>0$ for $u\neq 0$. A complex torus $\mathbb{C}^n/\Lambda$ is said to be ***polarizable*** if there exists a riemann form.

THEOREM 6.7. *The complex torus $\mathbb{C}^n/\Lambda$ is projective if and only if it is polarizable.*

PROOF. See Mumford 1970, Chapter I, (or Murty 1993, 4.1, for the "if" part). Alternatively, one can apply the Kodaira embedding theorem (Voisin 2002, Th. 7.11, 7.2.2). □

Thus, by Chow's theorem (3.11), a polarizable complex torus is a projective algebraic variety, and holomorphic maps of polarizable complex tori are regular. Conversely, it is easy to see that the complex manifold associated with an abelian variety is a complex torus: let $\mathrm{Tgt}_0 A$ be the tangent space to A at 0; then the exponential map $\mathrm{Tgt}_0 A\to A(\mathbb{C})$ is a surjective homomorphism of Lie groups with kernel a lattice Λ, which induces an isomorphism $(\mathrm{Tgt}_0 A)/\Lambda\cong A(\mathbb{C})$ of complex manifolds (Mumford 1970, p2).

For a complex torus $M=\mathbb{C}^n/\Lambda$, the isomorphism $\Lambda\otimes_{\mathbb{Z}}\mathbb{R}\cong\mathbb{C}^n$ endows $\Lambda\otimes_{\mathbb{Z}}\mathbb{R}$ with a complex structure, and hence endows $\Lambda\cong H_1(M,\mathbb{Z})$ with an integral hodge structure of weight -1. Note that a riemann form for M is nothing but a polarization of the integral hodge structure Λ.

THEOREM 6.8 (Riemann's theorem). [14] *The functor $A\mapsto H_1(A,\mathbb{Z})$ is an equivalence from the category AV of abelian varieties over $\mathbb{C}$ to the category of polarizable integral hodge structures of type $(-1,0),(0,-1)$.*

PROOF. We have functors

$$\mathsf{AV}\xrightarrow{A\mapsto A^{\mathrm{an}}}\{\text{category of polarizable complex tori}\}$$
$$\xrightarrow{M\mapsto H_1(M,\mathbb{Z})}\{\text{category of polarizable integral hodge structures of type}(-1,0),(0,-1)\}.$$

The first is fully faithful by Chow's theorem (3.11), and it is essentially surjective by Theorem 6.7; the second is fully faithful by Proposition 6.5, and it is obviously essentially surjective. □

Let AV^0 be the category whose objects are abelian varieties over $\mathbb{C}$ and whose morphisms are

$$\mathrm{Hom}_{\mathsf{AV}^0}(A,B)=\mathrm{Hom}_{\mathsf{AV}}(A,B)\otimes\mathbb{Q}.$$

COROLLARY 6.9. *The functor $A\mapsto H_1(A,\mathbb{Q})$ is an equivalence from the category AV^0 to the category of polarizable rational hodge structures of type $(-1,0),(0,-1)$.*

PROOF. Immediate consequence of the theorem. □

[14] In fact, it should be called the "theorem of Riemann, Frobenius, Weierstrass, Poincaré, Lefschetz, et al." (see Shafarevich 1994, Historical Sketch, 5), but "Riemann's theorem" is shorter.

REMARK 6.10. Recall that in the dictionary between complex structures J on a real vector space V and hodge structures of type $(-1,0),(0,-1)$,

$$(V,J)\cong V(\mathbb{C})/V^{-1,0}=V(\mathbb{C})/F^{0}.$$

Since the hodge structure on $H_1(A,\mathbb{R})$ is defined by the isomorphism $\mathrm{Tgt}_0(A)\cong H_1(A,\mathbb{R})$, we see that

$$\mathrm{Tgt}_0(A)\cong H_1(A,\mathbb{C})/F^{0} \tag{34}$$

(isomorphism of complex vector spaces).

A modular description of the points of the Siegel variety. Let $\mathcal{M}_K$ be the set of triples $(A,s,\eta K)$ in which A is an abelian variety over $\mathbb{C}$, s is an alternating form on $H_1(A,\mathbb{Q})$ such that s or $-s$ is a polarization on $H_1(A,\mathbb{Q})$, and η is an isomorphism $V(\mathbb{A}_f)\to V_f(\mathbb{A}_f)$ sending ψ to a multiple of s by an element of $\mathbb{A}_f^{\times}$. An isomorphism from one triple $(A,s,\eta K)$ to a second $(A',s',\eta' K)$ is an isomorphism $A\to A'$ (as objects in AV^0) sending s to a multiple of s' by an element of $\mathbb{Q}^{\times}$ and ηK to $\eta' K$.

THEOREM 6.11. *The set* $\mathrm{Sh}_K(G,X)$ *classifies the triples* $(A,s,\eta K)$ *in* $\mathcal{M}_K$ *modulo isomorphism, i.e., there is a canonical bijection* $\mathcal{M}_K/\approx\;\to G(\mathbb{Q})\backslash X\times G(\mathbb{A}_f)/K$.

PROOF. Combine (6.9) with (6.3). □

7. Shimura varieties of hodge type

In this section, we examine one important generalization of Siegel modular varieties.

DEFINITION 7.1. A Shimura datum (G,X) is of ***hodge type*** if there exists a symplectic space (V,ψ) over $\mathbb{Q}$ and an injective homomorphism $\rho\colon G\hookrightarrow G(\psi)$ carrying X into $X(\psi)$. The Shimura variety $\mathrm{Sh}(G,X)$ is then said to be of ***hodge type***. Here $(G(\psi),X(\psi))$ denotes the Shimura datum defined by (V,ψ).

The composite of ρ with the character ν of $G(\psi)$ is a character of G, which we again denote by ν. Let $\mathbb{Q}(r)$ denote the vector space $\mathbb{Q}$ with G acting by $r\nu$, i.e., $g\cdot v=\nu(g)^r\cdot v$. For each $h\in X$, $(\mathbb{Q}(r),h\circ\nu)$ is a rational hodge structure of type $(-r,-r)$ (apply 6.2a), and so this notation is consistent with that in (2.6).

LEMMA 7.2. *There exist multilinear maps* $t_i\colon V\times\cdots\times V\to\mathbb{Q}(r_i)$, $1\le i\le n$, *such that* G *is the subgroup of* $G(\psi)$ *fixing the* t_i.

PROOF. According to Deligne 1982, 3.1, there exist tensors t_i in $V^{\otimes r_i}\otimes V^{\vee\otimes s_i}$ such that this is true. But ψ defines an isomorphism $V\cong V^{\vee}\otimes\mathbb{Q}(1))$, and so

$$V^{\otimes r_i}\otimes V^{\vee\otimes s_i}\cong V^{\vee\otimes(r_i+s_i)}\otimes\mathbb{Q}(r_i)\cong\mathrm{Hom}(V^{\otimes(r_i+s_i)},\mathbb{Q}(r_i)).\qquad\square$$

Let (G,X) be of hodge type. Choose an embedding of (G,X) into $(G(\psi),X(\psi))$ for some symplectic space (V,ψ) and multilinear maps $t_1,\ldots,t_n$ as in the lemma. Let $\mathcal{H}_K$ be the set of triples $((W,h),(s_i)_{0\le i\le n},\eta K)$ in which

- (W,h) is a rational hodge structure of type $(-1,0),(0,-1)$,
- $\pm s_0$ is a polarization for (W,h),
- $s_1,\ldots,s_n$ are multilinear maps $s_i\colon W\times\cdots\times W\to\mathbb{Q}(r_i)$, and

- ηK is a K-orbit of isomorphisms $V(\mathbb{A}_f) \to W(\mathbb{A}_f)$ sending ψ onto an $\mathbb{A}_f^\times$-multiple of s_0 and each t_i to s_i,

satisfying the following condition:

> (*) there exists an isomorphism $a\colon W \to V$ sending s_0 to a $\mathbb{Q}^\times$-multiple of ψ, s_i to t_i each $i \geq 1$, and h onto an element of X.

An isomorphism from one triple $(W, \ldots)$ to a second $(W', \ldots)$ is an isomorphism $(W, h) \to (W', h')$ of rational hodge structures sending s_0 to a $\mathbb{Q}^\times$-multiple of s_0', s_i to s_i' for $i > 0$, and ηK to $\eta' K$.

PROPOSITION 7.3. *The set* $\mathrm{Sh}_K(G,X)(\mathbb{C})$ *classifies the triples in* $\mathcal{H}_K$ *modulo isomorphism.*

PROOF. Choose an isomorphism $a\colon W \to V$ as in (*), and consider the pair $(ah, a\circ\eta)$. By assumption $ah \in X$ and $a \circ \eta$ is a symplectic similitude of $(V(\mathbb{A}_f), \psi)$ fixing the t_i, and so $(ah, a \circ \eta) \in X \times G(\mathbb{A}_f)$. The isomorphism a is determined up to composition with an element of $G(\mathbb{Q})$ and η is determined up to composition with an element of K. It follows that the class of $(ah, a\circ\eta)$ in $G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K$ is well-defined. The proof that $(W, \ldots) \mapsto [ah, a \circ \eta]_K$ gives a bijection from the set of isomorphism classes of triples in $\mathcal{H}_K$ onto $\mathrm{Sh}_K(G,X)(\mathbb{C})$ is now routine (cf. the proof of 6.3). □

Let $t\colon V \times \cdots \times V \to \mathbb{Q}(r)$ (m-copies of V) be a multilinear form fixed by G, i.e., such that

$$t(gv_1, \ldots, gv_m) = \nu(g)^r \cdot t(v_1, \ldots, v_m), \text{ for all } v_1, \ldots, v_m \in V, \quad g \in G(\mathbb{Q}).$$

For $h \in X$, this equation shows that t defines a morphism of hodge structures $(V,h)^{\otimes m} \to \mathbb{Q}(r)$. On comparing weights, we see that if t is nonzero, then $m = 2r$.

Now let A be an abelian variety over $\mathbb{C}$, and let $V = H_1(A, \mathbb{Q})$. Then (see 6.4)

$$H^m(A,\mathbb{Q}) \cong \mathrm{Hom}(\textstyle\bigwedge^m V, \mathbb{Q}).$$

We say that $t \in H^{2r}(A,\mathbb{Q})$ is a ***hodge tensor for*** A if the corresponding map

$$V^{\otimes 2r} \to \textstyle\bigwedge^{2r} V \to \mathbb{Q}(r)$$

is a morphism of hodge structures.

Let $(G,X) \hookrightarrow (G(\psi), X(\psi))$ and $t_1, \ldots, t_n$ be as above. Let $\mathcal{M}_K$ be the set of triples $(A, (s_i)_{0\leq i\leq n}, \eta K)$ in which

- A is a complex abelian variety,
- $\pm s_0$ is a polarization for the rational hodge structure $H_1(A,\mathbb{Q})$,
- $s_1, \ldots, s_n$ are hodge tensors for A or its powers, and
- ηK is a K-orbit of $\mathbb{A}_f$-linear isomorphisms $V(\mathbb{A}_f) \to V_f(A)$ sending ψ onto an $\mathbb{A}_f^\times$-multiple of s_0 and each t_i to s_i,

satisfying the following condition:

> (**) there exists an isomorphism $a\colon H_1(A,\mathbb{Q}) \to V$ sending s_0 to a $\mathbb{Q}^\times$-multiple of ψ, s_i to t_i each $i \geq 1$, and h to an element of X.

An isomorphism from one triple $(A, (s_i)_i, \eta K)$ to a second $(A', (s_i'), \eta' K)$ is an isomorphism $A \to A'$ (as objects of AV^0) sending s_0 to a multiple of s_0' by an element of $\mathbb{Q}^\times$, each s_i to s_i', and η to η' modulo K.

THEOREM 7.4. *The set* $\mathrm{Sh}_K(G,X)(\mathbb{C})$ *classifies the triples in* $\mathcal{M}_K$ *modulo isomorphism.*

PROOF. Combine Propositions 7.3 and 6.9. □

The problem with Theorem 7.4 is that it is difficult to check whether a triple satisfies the condition (**). In the next section, we show that when the hodge tensors are endomorphisms of the abelian variety, then it is sometimes possible to replace (**) by a simpler trace condition.

REMARK 7.5. When we write $A(\mathbb{C}) = \mathbb{C}^g/\Lambda$, then (see 6.4),

$$H^m(A,\mathbb{Q}) \cong \mathrm{Hom}(\textstyle\bigwedge^m \Lambda, \mathbb{Q})$$

Now $\Lambda \otimes \mathbb{C} \cong T \oplus \overline{T}$ where $T = \mathrm{Tgt}_0(A)$. Therefore,

$$H^m(A,\mathbb{C}) \cong \mathrm{Hom}(\textstyle\bigwedge^m(\Lambda\otimes\mathbb{C}),\mathbb{C}) \cong \mathrm{Hom}(\displaystyle\bigoplus_{p+q=m} \textstyle\bigwedge^p T \otimes \bigwedge^q \overline{T}, \mathbb{C}) \cong \displaystyle\bigoplus_{p+q=m} H^{p,q}$$

where

$$H^{p,q} = \mathrm{Hom}(\textstyle\bigwedge^p T \otimes \bigwedge^q \overline{T}, \mathbb{C}).$$

This rather ad hoc construction of the hodge structure on H^m does agree with the usual construction (2.5) — see Mumford 1970, Chapter I. A hodge tensor on A is an element of

$$H^{2r}(A,\mathbb{Q}) \cap H^{r,r} \quad \text{(intersection inside } H^{2r}(A,\mathbb{C})).$$

The Hodge conjecture predicts that all hodge tensors are the cohomology classes of algebraic cycles with $\mathbb{Q}$-coefficients. For $r = 1$, this is known even over $\mathbb{Z}$. The exponential sequence

$$0 \to \mathbb{Z} \longrightarrow \mathcal{O}_A \xrightarrow{z\mapsto \exp(2\pi i z)} \mathcal{O}_A^\times \to 0$$

gives a cohomology sequence

$$H^1(A,\mathcal{O}_A^\times) \to H^2(A,\mathbb{Z}) \to H^2(A,\mathcal{O}_A).$$

The cohomology group $H^1(A,\mathcal{O}_A^\times)$ classifies the divisors on A modulo linear equivalence, i.e., $\mathrm{Pic}(A) \cong H^1(A,\mathcal{O}_A^\times)$, and the first arrow maps a divisor to its cohomology class. A class in $H^2(A,\mathbb{Z})$ maps to zero in $H^2(A,\mathcal{O}_A) = H^{0,2}$ if and only if it maps to zero in its complex conjugate $H^{2,0}$. Therefore, we see that

$$\mathrm{Im}(\mathrm{Pic}(A)) = H^2(A,\mathbb{Z}) \cap H^{1,1}.$$

8. PEL Shimura varieties

Throughout this section, k is a field of characteristic zero. Bilinear forms are always nondegenerate.

Algebras with involution. By a ***k-algebra*** I mean a ring B containing k in its centre and finite dimensional over k. A k-algebra A is ***simple*** if it contains no two-sided ideals except 0 and A. For example, every matrix algebra $M_n(D)$ over a division algebra D is simple, and conversely, Wedderburn's theorem says that every simple algebra is of this form (CFT, IV 1.9). Up to isomorphism, a simple k-algebra has only one simple module (ibid, IV 1.15). For example, up to isomorphism, D^n is the only simple $M_n(D)$-module.

Let $B = B_1 \times \cdots \times B_n$ be a product of simple k-algebras (a ***semisimple k-algebra)***. A simple B_i-module M_i becomes a simple B-module when we let B

act through the quotient map $B \to B_i$. These are the only simple B-modules, and every B-module is a direct sum of simple modules. A B-module M defines a k-linear map

$$b \mapsto \mathrm{Tr}_k(b|M) \colon B \to k$$

which we call the ***trace map*** of M.

PROPOSITION 8.1. *Let B be a semisimple k-algebra. Two B-modules are isomorphic if and only if they have the same trace map.*

PROOF. Let $B_1, \ldots, B_n$ be the simple factors of B, and let M_i be a simple B_i-module. Then every B-module is isomorphic to a direct sum $\bigoplus_j r_j M_j$ with $r_j M_j$ the direct sum of r_j copies of M_i. We have to show that the trace map determines the multiplicities r_j. But for $e_i = (0, \ldots, 0, \underset{i}{1}, 0, \ldots)$,

$$\mathrm{Tr}_k(e_i|\textstyle\sum r_j M_j) = r_i \dim_k M_i\,. \qquad \square$$

REMARK 8.2. The lemma fails when k has characteristic p, because the trace map is identically zero on pM.

An ***involution*** of a k-algebra B is a k-linear map $b \mapsto b^* \colon B \to B$ such that $(ab)^* = b^* a^*$ and $b^{**} = b$. Note that then $1^* = 1$ and so $c^* = c$ for $c \in k$.

PROPOSITION 8.3. *Let k be an algebraically closed field, and let $(B, *)$ be a semisimple k-algebra with involution. Then $(B, *)$ is isomorphic to a product of pairs of the following types:*

(A): $M_n(k) \times M_n(k)$, $\quad (a, b)^* = (b^t, a^t)$;
(C): $M_n(k)$, $\quad b^* = b^t$;
(BD): $M_n(k)$, $\quad b^* = J \cdot b^t \cdot J^{-1}$, $\quad J = \left(\begin{smallmatrix} 0 & -I \\ I & 0 \end{smallmatrix}\right)$.

PROOF. The decomposition $B = B_1 \times \cdots \times B_r$ of B into a product of simple algebras B_i is unique up to the ordering of the factors (Farb and Dennis 1993, 1.13). Therefore, $*$ permutes the set of B_i, and B is a product of semisimple algebras with involution each of which is either (i) simple or (ii) the product of two simple algebras interchanged by $*$.

Let $(B, *)$ be as in (i). Then B is isomorphic to $M_n(k)$ for some n, and the Noether-Skolem theorem (CFT, 2.10) shows that $b^* = u \cdot b^t \cdot u^{-1}$ for some $u \in M_n(k)$. Then $b = b^{**} = (u^t u^{-1})^{-1} b (u^t u^{-1})$ for all $b \in B$, and so $u^t u^{-1}$ lies in the centre k of $M_n(k)$. Denote it by c, so that $u^t = cu$. Then $u = u^{tt} = c^2 u$, and so $c^2 = 1$. Therefore, $u^t = \pm u$, and u is either symmetric or skew-symmetric. Relative to a suitable basis, u is I or J, and so $(B, *)$ is of type (C) or (BD).

Let $(B, *)$ be as in (ii). Then $*$ is an isomorphism of the opposite of the first factor onto the second. The Noether-Skolem theorem then shows that $(B, *)$ is isomorphic to $M_n(k) \times M_n(k)^{\mathrm{opp}}$ with the involution $(a, b) \mapsto (b, a)$. Now use that $a \leftrightarrow a^t \colon M_n(k)^{\mathrm{opp}} \cong M_n(k)$ to see that $(B, *)$ is of type (A). $\square$

The following is a restatement of the proposition.

PROPOSITION 8.4. *Let $(B, *)$ and k be as in (8.3). If the only elements of the centre of B invariant under $*$ are those in k, then $(B, *)$ is isomorphic to one of the following:*

(A): $\mathrm{End}_k(W) \times \mathrm{End}_k(W^\vee)$, $\quad (a, b)^* = (b^t, a^t)$;
(C): $\mathrm{End}_k(W)$, *b^* the transpose of b with respect to a symmetric bilinear form on W;*

(BD): $\mathrm{End}_k(W)$, *b^* the transpose of b with respect to an alternating bilinear form on W.*

Symplectic modules and the associated algebraic groups. Let $(B,*)$ be a semisimple k-algebra with involution $*$, and let (V,ψ) be a ***symplectic*** $(B,*)$-***module***, i.e., a B-module V endowed with an alternating k-bilinear form $\psi\colon V\times V\to k$ such that

$$\psi(bu,v)=\psi(u,b^*v)\text{ for all }b\in B,u,v\in V. \tag{35}$$

Let F be the centre of B, and let F_0 be the subalgebra of invariants of $*$ in F. Assume that B and V are free over F and that for all k-homomorphisms $\rho\colon F_0\to k^{\mathrm{al}}$, $(B\otimes_{F_0,\rho}k^{\mathrm{al}},*)$ is of the same type (A), (C), or (BD). This will be the case, for example, if F is a field. Let G be the subgroup of $\mathrm{GL}(V)$ such that

$$G(\mathbb{Q})=\{g\in\mathrm{Aut}_B(V)\mid\psi(gu,gv)=\mu(g)\psi(u,v)\text{ some }\mu(g)\in k^\times\},$$

and let

$$G'=\mathrm{Ker}(\mu)\cap\mathrm{Ker}(\det).$$

EXAMPLE 8.5. (Type A.) Let F be $k\times k$ or a field of degree 2 over k, and let $B=\mathrm{End}_F(W)$ equipped with the involution $*$ defined by a hermitian form[15] $\phi\colon W\times W\to F$. Then $(B,*)$ is of type A. Let V_0 be an F-vector space, and let ψ_0 be a skew-hermitian form $V_0\times V_0\to F$. The bilinear form ψ on $V=W\otimes_F V_0$ defined by

$$\psi(w\otimes v,w'\otimes v')=\mathrm{Tr}_{F/k}(\phi(w,w')\psi_0(v,v')) \tag{36}$$

is alternating and satisfies (35): (V,ψ) is a symplectic $(B,*)$-module. Let $C=\mathrm{End}_B(V)$ (the centralizer of B in $\mathrm{End}_F(V)$). Then C is stable under the involution $*$ defined by ψ, and

$$G(k)=\{c\in C^\times\mid cc^*\in k^\times\} \tag{37}$$

$$G'(k)=\{c\in C^\times\mid cc^*=1,\quad\det(c)=1\}. \tag{38}$$

In fact, $C\cong\mathrm{End}_F(V_0)$ and $*$ is transposition with respect to ψ_0. Therefore, G is the group of symplectic similitudes of ψ_0 whose multiplier lies in k, and G' is the special unitary group of ψ_0.

Conversely, let $(B,*)$ be of type A, and assume

(a) the centre F of B is of degree 2 over k (so F is a field or $k\times k$);
(b) B is isomorphic to a matrix algebra over F (when F is a field, this just means that B is simple and split over F).

Then I claim that $(B,*,V,\psi)$ arises as in the last paragraph. To see this, let W be a simple B-module — condition (b) implies that $B\cong\mathrm{End}_F(W)$ and that $*$ is defined by a hermitian form $\phi\colon W\times W\to F$. As a B-module, V is a direct sum of copies of W, and so $V=W\otimes_F V_0$ for some F-vector space V_0. Choose an element f of $F\smallsetminus k$ whose square is in k. Then $f^*=-f$, and

$$\psi(v,v')=\mathrm{Tr}_{F/k}\left(f\Psi(v,v')\right)$$

for a unique hermitian form $\Psi\colon V\times V\to F$ (Deligne 1982, 4.6), which has the property that $\Psi(bv,v')=\Psi(v,b^*v')$. The form $(v,v')\mapsto f\Psi(v,v')$ is skew-hermitian,

[15]There is a unique involution of F fixing k, which we again denote $*$. To say that ϕ is hermitian means that it is F-linear in one variable and satisfies $\phi(w,v)=\phi(v,w)^*$.

and can be[16] written $f\Psi = \phi \otimes \psi_0$ with ψ_0 skew-hermitian on V_0. Now ψ, ϕ, ψ_0 are related by (36).

EXAMPLE 8.6. (Type C.) Let $B = \mathrm{End}_k(W)$ equipped with the involution $*$ defined by a symmetric bilinear form $\phi\colon W \times W \to k$. Let V_0 be a k-vector space, and let ψ_0 be an alternating form $V_0 \times V_0 \to k$. The bilinear form ψ on $V = W \otimes V_0$ defined by

$$\psi(w \otimes v, w' \otimes v') = \phi(w, w')\psi_0(v, v')$$

is alternating and satisfies (35). Let $C = \mathrm{End}_B(V)$. Then C is stable under the involution $*$ defined by ψ, and $G(k)$ and $G'(k)$ are described by the equations (37) and (38). In fact, $C \cong \mathrm{End}_k(V_0)$ and $*$ is transposition with respect to ψ_0. Therefore $G = \mathrm{GSp}(V_0, \psi_0)$ and $G' = \mathrm{Sp}(V_0, \psi_0)$. Every system $(B, *, V, \psi)$ with B simple and split over k arises in this way (cf. 8.5).

PROPOSITION 8.7. *For $(B, *)$ of type A or C, the group G is reductive (in particular, connected), and G' is semisimple and simply connected.*

PROOF. It suffices to prove this after extending the scalars to the algebraic closure of k. Then $(B, *, V, \psi)$ decomposes into quadruples of the types considered in Examples 8.5 and 8.6, and so the proposition follows from the calculations made there. □

REMARK 8.8. Assume B is simple, and let m be the ***reduced dimension*** of V,

$$m = \frac{\dim_F(V)}{[B\colon F]^{\frac{1}{2}}}.$$

In case (A), $G'_{\mathbb{Q}^{\mathrm{al}}} \approx (\mathrm{SL}_m)^{[F_0:\mathbb{Q}]}$ and in case (C), $G'_{\mathbb{Q}^{\mathrm{al}}} \approx (\mathrm{Sp}_m)^{[F_0:\mathbb{Q}]}$.

REMARK 8.9. In case (BD), the group G is not connected (G' is a special orthogonal group) although its identity component is reductive.

Algebras with positive involution. Let C be a semisimple $\mathbb{R}$-algebra with an involution $*$, and let V be a C-module. In the next proposition, by a ***hermitian form*** on V we mean a symmetric bilinear form $\psi\colon V \times V \to \mathbb{R}$ satisfying (35). Such a form is said to be ***positive definite*** if $\psi(v, v) > 0$ for all nonzero $v \in V$.

PROPOSITION 8.10. *Let C be a semisimple algebra over $\mathbb{R}$. The following conditions on an involution $*$ of C are equivalent:*

(a) *some faithful C-module admits a positive definite hermitian form;*
(b) *every C-module admits a positive definite hermitian form;*
(c) $\mathrm{Tr}_{C/\mathbb{R}}(c^*c) > 0$ *for all nonzero $c \in C$.*

PROOF. (a) $\implies$ (b). Let V be a faithful C-module. Then every C-module is a direct summand of a direct sum of copies of V (see p323). Hence, if V carries a positive definite hermitian form, then so does every C-module.

(b) $\implies$ (c). Let V be a C-module with a positive definite hermitian form $(\,|\,)$, and choose an orthonormal basis $e_1, \ldots, e_n$ for V. Then

$$\mathrm{Tr}_{\mathbb{R}}(c^*c|V) = \textstyle\sum_i (e_i|c^*ce_i) = \sum_i (ce_i|ce_i),$$

[16] Probably the easiest way to prove things like this is use the correspondence between involutions on algebras and (skew-)hermitian forms (up to scalars) — see Knus et al. 1998, I 4.2. The involution on $\mathrm{End}_F(V)$ defined by ψ stabilizes C and corresponds to a skew-hermitian form on V_0.

which is > 0 unless c acts as the zero map on V. On applying this remark with $V = C$, we obtain (c).

(c) $\implies$ (a). The condition (c) is that the hermitian form $(c, c') \mapsto \mathrm{Tr}_{C/\mathbb{R}}(c^* c')$ on C is positive definite. □

DEFINITION 8.11. An involution satisfying the equivalent conditions of (8.10) is said to be ***positive***.

PROPOSITION 8.12. *Let B be a semisimple $\mathbb{R}$-algebra with a positive involution $*$ of type A or C. Let (V, ψ) be a symplectic $(B, *)$-module, and let C be the centralizer of B in $\mathrm{End}_{\mathbb{R}}(V)$. Then there exists a homomorphism of $\mathbb{R}$-algebras $h\colon \mathbb{C} \to C$, unique up to conjugation by an element c of $C^\times$ with $cc^* = 1$, such that*

- *$h(\bar{z}) = h(z)^*$ and*
- *$u, v \mapsto \psi(u, h(i)v)$ is positive definite and symmetric.*

PROOF. To give an h satisfying the conditions amounts to giving an element J $(= h(i))$ of C such that

$$J^2 = -1, \quad \psi(Ju, Jv) = \psi(u, v), \quad \psi(v, Jv) > 0 \text{ if } v \neq 0. \tag{39}$$

Suppose first that $(B, *)$ is of type A. Then $(B, *, V, \psi)$ decomposes into systems arising as in (8.5). Thus, we may suppose $B = \mathrm{End}_F(W)$, $V = W \otimes V_0$, etc., as in (8.5). We then have to classify the $J \in C \cong \mathrm{End}_{\mathbb{C}}(V_0)$ satisfying (39) with ψ replaced by ψ_0. There exists a basis (e_j) for V_0 such that

$$(\psi_0(e_j, e_k))_{j,k} = \mathrm{diag}(i, \ldots, \underset{r}{i}, -i, \ldots, -i), \quad i = \sqrt{-1}.$$

Define J by $J(e_j) = -\psi_0(e_j, e_j)e_j$. Then J satisfies the required conditions, and it is uniquely determined up to conjugation by an element of the unitary group of ψ_0. This proves the result for type A, and type C is similar. (For more details, see Zink 1983, 3.1). □

REMARK 8.13. Let $(B, *)$ and (V, ψ) be as in the proposition. For an h satisfying the conditions of the proposition, define

$$t(b) = \mathrm{Tr}_{\mathbb{C}}(b|V/F_h^0 V), \quad b \in B.$$

Then, t is independent of the choice of h, and in fact depends only on the isomorphism class of (V, ψ) as a B-module. Conversely, (V, ψ) is determined up to B-isomorphism by its dimension and t. For example, if $V = W \otimes_{\mathbb{C}} V_0$, ϕ, ψ_0, etc. are as in the above proof, then

$$\mathrm{Tr}_k(b|V) = r \cdot \mathrm{Tr}_k(b|W),$$

and r and $\dim V_0$ determine (V_0, ψ_0) up to isomorphism. Since W and ϕ are determined (up to isomorphism) by the requirement that W be a simple B-module and ϕ be a hermitian form giving $*$ on B, this proves the claim for type A.

PEL data. Let B be a simple $\mathbb{Q}$-algebra with a positive involution $*$ (meaning that it becomes positive on $B \otimes_{\mathbb{Q}} \mathbb{R}$), and let (V, ψ) be a symplectic $(B, *)$-module. Throughout this subsection, we assume that $(B, *)$ is of type A or C.

PROPOSITION 8.14. *There is a unique $G(\mathbb{R})$-conjugacy class X of homomorphisms $h\colon \mathbb{S} \to G_{\mathbb{R}}$ such that each $h \in X$ defines a complex structure on $V(\mathbb{R})$ that is positive or negative for ψ. The pair (G, X) satisfies the conditions SV1–4.*

Proof. The first statement is an immediate consequence of (8.12). The composite of h with $G \hookrightarrow G(\psi)$ lies in $X(\psi)$, and therefore satisfies SV1, SV2, SV4. As h is nontrivial, SV3 follows from the fact that G^{ad} is simple. $\square$

Definition 8.15. The Shimura data arising in this way are called ***simple PEL data of type A or C***.

The simple refers to the fact that (for simplicity), we required B to be simple (which implies that G^{ad} is simple).

Remark 8.16. Let $b \in B$, and let t_b be the tensor $(x, y) \mapsto \psi(x, by)$ of V. An element g of $G(\psi)$ fixes t_b if and only if it commutes with b. Let $b_1, \ldots, b_s$ be a set of generators for B as a $\mathbb{Q}$-algebra. Then (G, X) is the Shimura datum of hodge type associated with the system $(V, \{\psi, t_{b_1}, \ldots, t_{b_s}\})$.

PEL Shimura varieties.

Theorem 8.17. *Let (G, X) be a simple PEL datum of type A or C associated with $(B, *, V, \psi)$ as in the last subsection, and let K be a compact open subgroup of $G(\mathbb{A}_f)$. Then $\mathrm{Sh}_K(G, X)(\mathbb{C})$ classifies the isomorphism classes of quadruples $(A, s, i, \eta K)$ in which*

- *A is a complex abelian variety,*
- *$\pm s$ is a polarization of the hodge structure $H_1(A, \mathbb{Q})$,*
- *i is a homomorphism $B \to \mathrm{End}(A) \otimes \mathbb{Q}$, and*
- *ηK is a K-orbit of $B \otimes \mathbb{A}_f$-linear isomorphisms $\eta\colon V(\mathbb{A}_f) \to H^1(A, \mathbb{Q}) \otimes \mathbb{A}_f$ sending ψ to an $\mathbb{A}_f^{\times}$-multiple of s,*

satisfying the following condition:

> *(**) there exists a B-linear isomorphism $a\colon H_1(A, \mathbb{Q}) \to V$ sending s to a $\mathbb{Q}^{\times}$-multiple of ψ.*

Proof. In view of the dictionary $b \leftrightarrow t_b$ between endomorphisms and tensors (8.16), Theorem 7.4 shows that $\mathrm{Sh}_K(G, X)(\mathbb{C})$ classifies the quadruples $(A, i, t, \eta K)$ with the additional condition that $ah \in X$, but ah defines a complex structure on $V(\mathbb{R})$ that is positive or negative for ψ, and so (8.14) shows that ah automatically lies in X. $\square$

Let (G, X) be the Shimura datum arising from $(B, *)$ and (V, ψ). For $h \in X$, we have a trace map

$$b \mapsto \mathrm{Tr}(b|V(\mathbb{C})/F_h^0)\colon B \to \mathbb{C}.$$

Since this map is independent of the choice of h in X, we denote it by Tr_X.

Remark 8.18. Consider a triple $(A, s, i, \eta K)$ as in the theorem. The existence of the isomorphism a in (**) implies that

(a) $s(bu, v) = s(u, b^* v)$, and
(b) $\mathrm{Tr}(i(b)|\mathrm{Tgt}_0 A) = \mathrm{Tr}_X(b)$ for all $b \in B \otimes \mathbb{C}$.

The first is obvious, because ψ has this property, and the second follows from the B-isomorphisms

$$\mathrm{Tgt}_0(A) \overset{(34)}{\cong} H_1(A, \mathbb{C})/F^0 \xrightarrow{a} V(\mathbb{C})/F_h^0.$$

We now divide the type A in two, depending on whether the reduced dimension of V is even or odd.

PROPOSITION 8.19. *For types Aeven and C, the condition (**) of Theorem 8.17 is implied by conditions (a) and (b) of (8.18).*

PROOF. Let $W = H_1(A, \mathbb{Q})$. We have to show that there exists a B-linear isomorphism $\alpha\colon W \to V$ sending s to a $\mathbb{Q}^\times$-multiple of ψ. The existence of η shows that W has the same dimension as V, and so there exists a $B \otimes_{\mathbb{Q}} \mathbb{Q}^{\mathrm{al}}$-isomorphism $\alpha\colon V(\mathbb{Q}^{\mathrm{al}}) \to W(\mathbb{Q}^{\mathrm{al}})$ sending t to a $\mathbb{Q}^{\mathrm{al}\times}$-multiple of ψ. For $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ write $\sigma\alpha = \alpha \circ a_\sigma$ with $a_\sigma \in G(\mathbb{Q}^{\mathrm{al}})$. Then $\sigma \mapsto a_\sigma$ is a one-cocycle. If its class in $H^1(\mathbb{Q}, G)$ is trivial, say, $a_\sigma = a^{-1} \cdot \sigma a$, then $\alpha \circ a^{-1}$ is fixed by all $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$, and is therefore defined over $\mathbb{Q}$.

Thus, it remains to show that the class of (a_σ) in $H^1(\mathbb{Q}, G)$ is trivial. The existence of η shows that the image of the class in $H^1(\mathbb{Q}_\ell, G)$ is trivial for all finite primes ℓ, and (8.13) shows that its image in $H^1(\mathbb{R}, G)$ is trivial, and so the statement follows from the next two lemmas. □

LEMMA 8.20. *Let G be a reductive group with simply connected derived group, and let $T = G/G^{\mathrm{der}}$. If $H^1(\mathbb{Q}, T) \to \prod_{l \leq \infty} H^1(\mathbb{Q}_l, T)$ is injective, then an element of $H^1(\mathbb{Q}, G)$ that becomes trivial in $H^1(\mathbb{Q}_l, G)$ for all l is itself trivial.*

PROOF. Because G^{der} is simply connected, $H^1(\mathbb{Q}_l, G^{\mathrm{der}}) = 0$ for $l \neq \infty$ and $H^1(\mathbb{Q}, G^{\mathrm{der}}) \to H^1(\mathbb{R}, G^{\mathrm{der}})$ is injective (5.19). Using this, we obtain a commutative diagram with exact rows

$$\begin{tikzcd}
 & T(\mathbb{Q}) \arrow[r] \arrow[d] & H^1(\mathbb{Q}, G^{\mathrm{der}}) \arrow[r] \arrow[d, "\text{injective}"'] & H^1(\mathbb{Q}, G) \arrow[r] \arrow[d] & H^1(\mathbb{Q}, T) \arrow[d, "\text{injective}"'] \\
G(\mathbb{R}) \arrow[r] & T(\mathbb{R}) \arrow[r] & H^1(\mathbb{R}, G^{\mathrm{der}}) \arrow[r] & \prod_l H^1(\mathbb{Q}_l, G) \arrow[r] & \prod_l H^1(\mathbb{Q}_l, T).
\end{tikzcd}$$

If an element c of $H^1(\mathbb{Q}, G)$ becomes trivial in all $H^1(\mathbb{Q}_l, G)$, then a diagram chase shows that it arises from an element c' of $H^1(\mathbb{Q}, G^{\mathrm{der}})$ whose image c'_∞ in $H^1(\mathbb{R}, G^{\mathrm{der}})$ maps to the trivial element in $H^1(\mathbb{R}, G)$. The image of $G(\mathbb{R})$ in $T(\mathbb{R})$ contains $T(\mathbb{R})^+$ (see 5.1), and the real approximation theorem (5.4) shows that $T(\mathbb{Q}) \cdot T(\mathbb{R})^+ = T(\mathbb{R})$. Therefore, there exists a $t \in T(\mathbb{Q})$ whose image in $H^1(\mathbb{R}, G^{\mathrm{der}})$ is c'_∞. Then $t \mapsto c'$ in $H^1(\mathbb{Q}, G^{\mathrm{der}})$, which shows that c is trivial. □

LEMMA 8.21. *Let (G, X) be a simple PEL Shimura datum of type Aeven or C, and let $T = G/G^{\mathrm{der}}$. Then $H^1(\mathbb{Q}, T) \to \prod_{l \leq \infty} H^1(\mathbb{Q}_l, T)$ is injective.*

PROOF. For G of type Aeven, $T = \mathrm{Ker}((\mathbb{G}_m)_F \xrightarrow{\mathrm{Nm}_{F/k}} (\mathbb{G}_m)_{F_0}) \times \mathbb{G}_m$. The group $H^1(\mathbb{Q}, \mathbb{G}_m) = 0$, and the map on H^1's of the first factor is

$$F_0^\times / \operatorname{Nm} F^\times \to \textstyle\prod_v F_{0v}^\times / \operatorname{Nm} F_v^\times.$$

This is injective (CFT, VIII 1.4).

For G of type C, $T = \mathbb{G}_m$, and so $H^1(\mathbb{Q}, T) = 0$. □

PEL modular varieties. Let B be a semisimple algebra over $\mathbb{Q}$ with a positive involution $*$, and let (V, ψ) be a symplectic $(B, *)$-module. Let K be a compact open subgroup of $G(\mathbb{A}_f)$. There exists an algebraic variety M_K over $\mathbb{C}$ classifying the isomorphism classes of quadruples $(A, s, i, \eta K)$ satisfying (a) and (b) of (8.18) (but not necessarily condition (**)), which is called the ***PEL modular variety attached to*** $(B, *, V, \psi)$. In the simple cases (Aeven) and (C), Proposition 8.17

shows that M_K coincides with $\mathrm{Sh}_K(G, X)$, but in general it is a finite disjoint union of Shimura varieties.

Notes. The theory of Shimura varieties of PEL-type is worked out in detail in several papers of Shimura, for example, Shimura 1963, but in a language somewhat different from ours. The above account follows Deligne 1971*c*, §§5,6. See also Zink 1983 and Kottwitz 1992, §§1–4.

9. General Shimura varieties

Abelian motives. Let $\mathsf{Hod}(\mathbb{Q})$ be the category of polarizable rational hodge structures. It is an abelian subcategory of the category of all rational hodge structures closed under the formation of tensor products and duals.

Let V be a variety over $\mathbb{C}$ whose connected components are abelian varieties, say $V = \bigsqcup V_i$ with V_i an abelian variety. Recall that for manifolds M_1 and M_2,

$$H^r(M_1 \sqcup M_2, \mathbb{Q}) \cong H^r(M_1, \mathbb{Q}) \oplus H^r(M_2, \mathbb{Q}).$$

For each connected component V° of V,

$$H^*(V^\circ, \mathbb{Q}) \cong \bigwedge H^1(V^\circ, \mathbb{Q}) \cong \mathrm{Hom}_{\mathbb{Q}}(\bigwedge H_1(V^\circ, \mathbb{Q}), \mathbb{Q})$$

(see 6.4). Therefore, $H^*(V, \mathbb{Q})$ acquires a polarizable hodge structure from that on $H_1(V, \mathbb{Q})$. We write $H^*(V, \mathbb{Q})(m)$ for the hodge structure $H^*(V, \mathbb{Q}) \otimes \mathbb{Q}(m)$ (see 2.6).

Let (W, h) be a rational hodge structure. An endomorphism e of (W, h) is an ***idempotent*** if $e^2 = e$. Then

$$(W, h) = \mathrm{Im}(e) \oplus \mathrm{Im}(1 - e)$$

(direct sum of rational hodge structures).

An ***abelian motive*** over $\mathbb{C}$ is a triple (V, e, m) in which V is a variety over $\mathbb{C}$ whose connected components are abelian varieties, e is an idempotent in $\mathrm{End}(H^*(V, \mathbb{Q}))$, and $m \in \mathbb{Z}$. For example, let A be an abelian variety; then the projection

$$H^*(A, \mathbb{Q}) \to H^i(A, \mathbb{Q}) \subset H^*(A, \mathbb{Q})$$

is an idempotent e^i, and we denote $(A, e^i, 0)$ by $h^i(A)$.

Define $\mathrm{Hom}((V, e, m), (V', e', m'))$ to be the set of maps $H^*(V, \mathbb{Q}) \to H^*(V', \mathbb{Q})$ of the form $e' \circ f \circ e$ with f a homomorphism $H^*(V, \mathbb{Q}) \to H^*(V', \mathbb{Q})$ of degree $d = m' - m$. Moreover, define

$$\begin{aligned}
(V, e, m) \oplus (V', e', m) &= (V \sqcup V', e \oplus e', m)\\
(V, e, m) \otimes (V', e', m) &= (V \times V', e \otimes e', m + m')\\
(V, e, m)^\vee &= (V, e^t, d - m) \text{ if } V \text{ is purely } d\text{-dimensional.}
\end{aligned}$$

For an abelian motive (V, e, m) over $\mathbb{C}$, let $H(V, e, m) = eH^*(V, \mathbb{Q})(m)$. Then $(V, e, m) \mapsto H(V, e, m)$ is a functor from the category of abelian motives AM to $\mathsf{Hod}(\mathbb{Q})$ commuting with $\oplus$, $\otimes$, and $^\vee$. We say that a rational hodge structure is ***abelian*** if it is in the essential image of this functor, i.e., if it is isomorphic to $H(V, e, m)$ for some abelian motive (V, e, m). Every abelian hodge structure is polarizable.

PROPOSITION 9.1. *Let* $\mathsf{Hod}^{\mathrm{ab}}(\mathbb{Q})$ *be the full subcategory of* $\mathsf{Hod}(\mathbb{Q})$ *of abelian hodge structures. Then* $\mathsf{Hod}^{\mathrm{ab}}(\mathbb{Q})$ *is the smallest strictly full subcategory of* $\mathsf{Hod}(\mathbb{Q})$ *containing* $H_1(A,\mathbb{Q})$ *for each abelian variety* A *and closed under the formation of direct sums, subquotients, duals, and tensor products; moreover,* $H\colon \mathsf{AM}\to\mathsf{Hod}^{\mathrm{ab}}(\mathbb{Q})$ *is an equivalence of categories.*

PROOF. Straightforward from the definitions. □

For a description of the essential image of H, see Milne 1994, 1.27.

Shimura varieties of abelian type. Recall (§6) that a symplectic space (V,ψ) over $\mathbb{Q}$ defines a connected Shimura datum $(S(\psi),X(\psi)^+)$ with $S(\psi)=\mathrm{Sp}(\psi)$ and $X(\psi)^+$ the set of complex structures on $V(\mathbb{R}),\psi)$.

DEFINITION 9.2. (a) A connected Shimura datum (H,X^+) with H simple is of ***primitive abelian type*** if there exists a symplectic space (V,ψ) and an injective homomorphism $H\to S(\psi)$ carrying X^+ into $X(\psi)^+$.

(b) A connected Shimura datum (H,X^+) is of ***abelian type*** if there exist pairs (H_i,X_i^+) of primitive abelian type and an isogeny $\prod_i H_i\to H$ carrying $\prod_i X_i^+$ into X.

(b) A Shimura datum (G,X) is of ***abelian type*** if (G^{der},X^+) is of abelian type.

(c) The (connected) Shimura variety attached to a (connected) Shimura datum of abelian type is said to be of ***abelian type***.

PROPOSITION 9.3. *Let* (G,X) *be a Shimura datum, and assume*

(a) *the weight* w_X *is rational SV4 and* $Z(G)^\circ$ *splits over a CM-field SV6, and*
(b) *there exists a homomorphism* $\nu\colon G\to\mathbb{G}_m$ *such that* $\nu\circ w_X=-2$.

If G *is of abelian type, then* $(V,h\circ\rho)$ *is an abelian hodge structure for all representations* (V,ρ) *of* G *and all* $h\in X$*; conversely, if there exists a faithful representation* ρ *of* G *such that* $(V,h\circ\rho)$ *is an abelian hodge structure for all* h*, then* (G,X) *is of abelian type.*

PROOF. See Milne 1994, 3.12. □

Let (G,X) be a Shimura datum of abelian type satisfying (a) and (b) of the proposition, and let $\rho\colon G\to\mathrm{GL}(V)$ be a faithful representation of G. Assume that there exists a pairing $\psi\colon V\times V\to\mathbb{Q}$ such that

(a) $g\psi=\nu(g)^m\psi$ for all $g\in G$,
(b) ψ is a polarization of $(V,h\circ\rho)$ for all $h\in X$.

There exist multilinear maps $t_i\colon V\times\cdots\times V\to\mathbb{Q}(r_i)$, $1\le i\le n$, such that G is the subgroup of $\mathrm{GL}(V)$ whose elements satisfy (a) and fix $t_1,\dots t_n$ (cf. 7.2).

THEOREM 9.4. *With the above notations,* $\mathrm{Sh}(G,X)$ *classifies the isomorphism classes of triples* $(A,(s_i)_{0\le i\le n},\eta K)$ *in which*

- A *is an abelian motive,*
- $\pm s_0$ *is a polarization for the rational hodge structure* $H(A)$,
- $s_1,\dots,s_n$ *are tensors for* A, *and*
- ηK *is a* K*-orbit of* $\mathbb{A}_f$*-linear isomorphisms* $V(\mathbb{A}_f)\to V_f(A)$ *sending* ψ *to an* $\mathbb{A}_f^\times$*-multiple of* s_0 *and each* t_i *to* s_i,

satisfying the following condition:

(**) there exists an isomorphism $a\colon H(A) \to V$ *sending* s_0 *to a* $\mathbb{Q}^\times$*-multiple of* ψ*, each* s_i *to* t_i*, and* h *onto an element of* X*.*

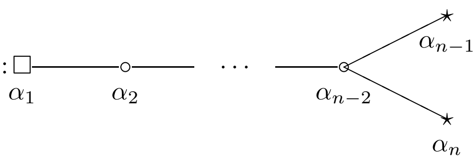
PROOF. With A replaced by a hodge structure, this can be proved by an elementary argument (cf. 6.3, 7.3), but (9.3) shows that the hodge structures arising are abelian, and so can be replaced by abelian motives (9.1). For more details, see Milne 1994, Theorem 3.31. $\square$

Classification of Shimura varieties of abelian type. Deligne (1979) classifies the connected Shimura data of abelian type. Let (G, X^+) be a connected Shimura datum with G simple. If G^{ad} is of type A, B, or C, then (G, X^+) is of abelian type. If G^{ad} is of type E_6 or E_7, then (G, X^+) is not of abelian type. If G^{ad} is of type D, (G, X^+) may or may not be of abelian type. There are two problems that may arise.

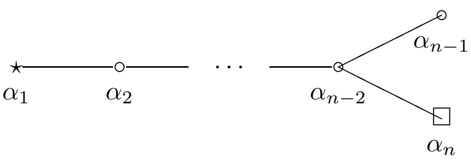
(a) Let G be the universal covering group of G^{ad}. There may exist homomorphisms $(G, X^+) \to (S(\psi), X(\psi)^+)$ but no injective such homomorphism, i.e., there may be a nonzero finite algebraic subgroup $N \subset G$ that is in the kernel of all homomorphisms $G \to S(\psi)$ sending X^+ into $X(\psi)^+$. Then $(G/N', X^+)$ is of abelian type for all $N' \supset N$, but (G, X^+) is not of abelian type.

(b) There may not exist a homomorphism $G \to S(\psi)$ at all.

This last problem arises for the following reason. Even when G^{ad} is $\mathbb{Q}$-simple, it may decompose into a product of simple group $G^{\mathrm{ad}}_{\mathbb{R}} = G_1 \times \cdots \times G_r$ over $\mathbb{R}$. For each i, G_i has a dynkin diagram of the shape shown below:

$D_n(1)$: 

$(n \geq 4)$

$D_n(n)$: 

$(n \geq 4)$

$D_n(n-1)$: Same as $D_n(n)$ by with α_{n-1} and α_n interchanged (rotation about the horizontal axis).

Nodes marked by squares are special (p278), and nodes marked by stars correspond to symplectic representations. The number in parenthesis indicates the position of the special node. As is explained in §1, the projection of X^+ to a conjugacy class of homomorphisms $\mathbb{S} \to G_i$ corresponds to a node marked with a $\square$. Since X^+ is defined over $\mathbb{R}$, the nodes can be chosen independently for each i. On the other hand, the representations $G_{i\mathbb{R}} \to S(\psi)_{\mathbb{R}}$ correspond to nodes marked with a $*$. Note that the $*$ has to be at the opposite end of the diagram from the $\square$. In order for a family of representations $G_{i\mathbb{R}} \to S(\psi)_{\mathbb{R}}$, $1 \leq i \leq r$, to arise from a symplectic representation over $\mathbb{Q}$, the $*$'s must be all in the same position since a galois group must permute the dynkin diagrams of the G_i. Clearly, this is impossible if the $\square$'s occur at different ends. (See Deligne 1979, 2.3, for more details.)

Shimura varieties not of abelian type. It is hoped (Deligne 1979, p248) that all Shimura varieties with rational weight classify isomorphism classes of motives with additional structure, but this is not known for those not of abelian type. More precisely, from the choice of a rational representation $\rho\colon G \to \mathrm{GL}(V)$, we obtain a family of hodge structures $h \circ \rho_{\mathbb{R}}$ on V indexed by X. When the weight of (G,X) is defined over $\mathbb{Q}$, it is hoped that these hodge structures always occur (in a natural way) in the cohomology of algebraic varieties. When the weight of (G,X) is not defined over $\mathbb{Q}$ they obviously can not.

Example: simple Shimura varieties of type A_1. Let (G,X) be the Shimura datum attached to a B be a quaternion algebra over a totally real field F, as in (5.24). With the notations of that example,

$$G(\mathbb{R}) \approx \prod\nolimits_{v\in I_c} \mathbb{H}^{\times} \times \prod\nolimits_{v\in I_{nc}} \mathrm{GL}_2(\mathbb{R}).$$

(a) If $B = M_2(F)$, then (G,X) is of PEL-type, and $\mathrm{Sh}_K(G,X)$ classifies isomorphism classes of quadruples $(A,i,t,\eta K)$ in which A is an abelian variety of dimension $d = [F\colon \mathbb{Q}]$ and i is a homomorphism homomorphism $i\colon F \to \mathrm{End}(A)\otimes\mathbb{Q}$. These Shimura varieties are called ***Hilbert*** (or ***Hilbert-Blumenthal***) varieties, and whole books have been written about them.

(b) If B is a division algebra, but $I_c = \emptyset$, then (G,X) is again of PEL-type, and $\mathrm{Sh}_K(G,X)$ classifies isomorphism classes of quadruples $(A,i,t,\eta K)$ in which A is an abelian variety of dimension $2[F\colon\mathbb{Q}]$ and i is a homomorphism $i\colon B \to \mathrm{End}(A)\otimes\mathbb{Q}$. In this case, the varieties are projective. These varieties have also been extensively studied.

(c) If B is a division algebra and $I_c \neq \emptyset$, then (G,X) is of abelian type, but the weight is not defined over $\mathbb{Q}$. Over $\mathbb{R}$, the weight map w_X sends $a \in \mathbb{R}$ to the element of $(F\otimes\mathbb{R})^{\times} \cong \prod_{v\colon F\to\mathbb{R}} \mathbb{R}$ with component 1 for $v \in I_c$ and component a for $v \in I_{nc}$. Let T be the torus over $\mathbb{Q}$ with $T(\mathbb{Q}) = F^{\times}$. Then $w_X\colon \mathbb{G}_m \to T_{\mathbb{R}}$ is defined over the subfield L of $\overline{\mathbb{Q}}$ whose fixed group is the subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ stabilizing $I_c \subset I_c \sqcup I_{nc}$. On choosing a rational representation of G, we find that $\mathrm{Sh}_K(G,X)$ classifies certain isomorphism classes of hodge structures with additional structure, but the hodge structures are not motivic — they do not arise in the cohomology of algebraic varieties (they are not rational hodge structures).

10. Complex multiplication: the Shimura-Taniyama formula

Where we are headed. Let V be a variety over $\mathbb{Q}$. For any $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ and $P \in V(\mathbb{Q}^{\mathrm{al}})$, the point $\sigma P \in V(\mathbb{Q}^{\mathrm{al}})$. For example, if V is the subvariety of $\mathbb{A}^n$ defined by equations

$$f(X_1,\ldots,X_n) = 0, \quad f \in \mathbb{Q}[X_1,\ldots,X_n],$$

then

$$f(a_1,\ldots,a_n) = 0 \Longrightarrow f(\sigma a_1,\ldots,\sigma a_n) = 0$$

(apply σ to the first equality). Therefore, if we have a variety V over $\mathbb{Q}^{\mathrm{al}}$ that we suspect is actually defined over $\mathbb{Q}$, then we should be able to describe an action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ on its points $V(\mathbb{Q}^{\mathrm{al}})$.

Let E be a number field contained in $\mathbb{C}$, and let $\mathrm{Aut}(\mathbb{C}/E)$ denote the group of automorphisms of $\mathbb{C}$ (as an abstract field) fixing the elements of E. Then a similar remark applies: if a variety V over $\mathbb{C}$ is defined by equations with coefficients in E, then $\mathrm{Aut}(\mathbb{C}/E)$ will act on $V(\mathbb{C})$. Now, I claim that all Shimura varieties are defined (in a natural way) over specific number fields, and so I should be able to describe an action of a big subgroup of $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ on their points. If, for example, the Shimura variety is of hodge type, then there is a set $\mathcal{M}_K$ whose elements are abelian varieties plus additional data and a map

$$(A, \ldots) \mapsto P(A, \ldots)\colon \mathcal{M}_K \to \mathrm{Sh}_K(G, X)(\mathbb{C})$$

whose fibres are the isomorphism classes in $\mathcal{M}_K$. On applying $\sigma \in \mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ to the coefficients of the polynomials defining $A, \ldots$, we get a new triple $(\sigma A, \ldots)$ which may or may not lie in $\mathcal{M}_K$. When it does we define ${}^{\sigma}P(A, \ldots)$ to be $P(\sigma A, \ldots)$. Our task will be to show that, for some specific field E, this does give an action of $\mathrm{Aut}(\mathbb{C}/E)$ on $\mathrm{Sh}_K(G, X)$ and that the action does arise from a model of $\mathrm{Sh}_K(G, X)$ over E.

For example, for $P \in \Gamma(1)\backslash\mathcal{H}_1$, ${}^{\sigma}P$ is the point such that $j({}^{\sigma}P) = \sigma\,(j(P))$. If j were a polynomial with coefficients in $\mathbb{Z}$ (rather than a power series with coefficients in $\mathbb{Z}$), we would have $j(\sigma P) = \sigma j(P)$ with the obvious meaning of σP, but this is definitely false (if σ is not complex conjugation, then it is not continuous, nor even measurable).

You may complain that we fail to explicitly describe the action of $\mathrm{Aut}(\mathbb{C}/E)$ on $\mathrm{Sh}(G, X)(\mathbb{C})$, but I contend that there can not exist a completely explicit description of the action. What are the elements of $\mathrm{Aut}(\mathbb{C}/E)$? To construct them, we can choose a transcendence basis B for $\mathbb{C}$ over E, choose a permutation of the elements of B, and extend the resulting automorphism of $\mathbb{Q}(B)$ to its algebraic closure $\mathbb{C}$. But proving the existence of transcendence bases requires the axiom of choice (e.g., FT, 8.13), and so we can have no explicit description of, or way of naming, the elements of $\mathrm{Aut}(\mathbb{C}/E)$, and hence no completely explicit description of the action is possible.

However, all is not lost. Abelian class field theory names the elements of $\mathrm{Gal}(E^{\mathrm{ab}}/E)$, where E^{ab} is a maximal abelian extension of E. Thus, if we suspect that a point P has coordinates in E^{ab}, the action of $\mathrm{Aut}(\mathbb{C}/E)$ on it will factor through $\mathrm{Gal}(E^{\mathrm{ab}}/E)$, and we may hope to be able to describe the action of $\mathrm{Aut}(\mathbb{C}/E)$ explicitly. This the theory of complex multiplication allows us to do for certain special points P.

Review of abelian varieties. The theory of abelian varieties is very similar to that of elliptic curves — just replace E with A, 1 with g (the dimension of A), and, whenever E occurs twice, replace one copy with the dual $A^{\vee}$ of A.

Thus, for any m not divisible by the characteristic of the ground field k,

$$A(k^{\mathrm{al}})_m \approx (\mathbb{Z}/m\mathbb{Z})^{2g}. \tag{40}$$

Here $A(k^{\mathrm{al}})_m$ consists of the elements of $A(k^{\mathrm{al}})$ killed by m. Hence, for $\ell \neq \mathrm{char}(k)$,

$$T_\ell A \overset{\mathrm{df}}{=} \varprojlim A(k^{\mathrm{al}})_{\ell^n}$$

is a free $\mathbb{Z}_\ell$-module of rank $2g$, and

$$V_\ell(A) \overset{\mathrm{df}}{=} T_\ell A \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$$

is a $\mathbb{Q}_\ell$-vector space of dimension $2g$. In characteristic zero, we set

$$T_f A = \textstyle\prod T_\ell A = \varprojlim_m A(k^{\mathrm{al}})_m,$$

$$V_f A = T_f \otimes_{\mathbb{Z}} \mathbb{Q} = \textstyle\prod (V_\ell A\colon T_\ell A) \text{ (restricted topological product).}$$

They are, respectively, a free $\hat{\mathbb{Z}}$-module of rank $2g$ and a free $\mathbb{A}_f$-module of rank $2g$. The galois group $\mathrm{Gal}(k^{\mathrm{al}}/k)$ acts continuously on these modules.

For an endomorphism a of an abelian variety A, there is a unique monic polynomial $P_a(T)$ with integer coefficients (the ***characteristic polynomial of*** a) such that $|P_a(n)| = \deg(a-n)$ for all $n \in \mathbb{Z}$. Moreover, P_a is the characteristic polynomial of a acting on $V_\ell A$ ($\ell \neq \mathrm{char}(k)$).

For an abelian variety A over a field k, the tangent space $\mathrm{Tgt}_0(A)$ to A at 0 is a vector space over k of dimension g. As we noted in §6, when $k = \mathbb{C}$, the exponential map defines a surjective homomorphism $\mathrm{Tgt}_0(A) \to A(\mathbb{C})$ whose kernel is a lattice Λ in $\mathrm{Tgt}_0(A)$. Thus $A(\mathbb{C})_m \cong \frac{1}{m}\Lambda/\Lambda \cong \Lambda/m\Lambda$, and

$$T_\ell A \cong \Lambda \otimes_{\mathbb{Z}} \mathbb{Z}_\ell, \quad V_\ell A \cong \Lambda \otimes_{\mathbb{Z}} \mathbb{Q}_\ell, \quad T_f A = \Lambda \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}, \quad V_f A = \Lambda \otimes_{\mathbb{Z}} \mathbb{A}_f. \tag{41}$$

An endomorphism a of A defines a $\mathbb{C}$-linear endomorphism $(da)_0 = \alpha$ of $\mathrm{Tgt}_0(A)$ such that $\alpha(\Lambda) \subset \Lambda$ (see 6.5), and $P_a(T)$ is the characteristic polynomial of α on Λ.

For abelian varieties A, B, $\mathrm{Hom}(A,B)$ is a torsion free $\mathbb{Z}$-module of finite rank. We let $\mathsf{AV}(k)$ denote the category of abelian varieties and homomorphisms over k and $\mathsf{AV}^0(k)$ the category with the same objects but with

$$\mathrm{Hom}_{\mathsf{AV}^0(k)}(A,B) = \mathrm{Hom}^0(A,B) = \mathrm{Hom}_{\mathsf{AV}(k)}(A,B) \otimes \mathbb{Q}.$$

An ***isogeny*** of abelian varieties is a surjective homomorphism with finite kernel. A homomorphism of abelian varieties is an isogeny if and only if it becomes an isomorphism in the category AV^0. Two abelian varieties are said to be ***isogenous*** if there is an isogeny from one to the other — this is an equivalence relation.

An abelian variety A over a field k is ***simple*** if it contains no nonzero proper abelian subvariety. Every abelian variety is isogenous to a product of simple abelian varieties. If A and B are simple, then every nonzero homomorphism from A to B is an isogeny. It follows that $\mathrm{End}^0(A)$ is a division algebra when A is simple and a semisimple algebra in general.

NOTES. For a detailed account of abelian varieties over algebraically closed fields, see Mumford 1970, and for a summary over arbitrary fields, see Milne 1986.

CM fields. A number field E is a ***CM*** (or ***complex multiplication***) ***field*** if it is a quadratic totally imaginary extension of a totally real field F. Let $a \mapsto a^*$ denote the nontrivial automorphism of E fixing F. Then $\rho(a^*) = \overline{\rho(a)}$ for every $\rho\colon E \hookrightarrow \mathbb{C}$. We have the following picture:

$$\begin{array}{ccc} E \otimes_{\mathbb{Q}} \mathbb{R} & \approx & \mathbb{C} \times \cdots \times \mathbb{C} \\ | & & | \\ F \otimes_{\mathbb{Q}} \mathbb{R} & \approx & \mathbb{R} \times \cdots \times \mathbb{R} \end{array} \tag{42}$$

The involution $*$ is positive (in the sense of 8.11), because we can compute $\mathrm{Tr}_{E\otimes_{\mathbb{Q}}\mathbb{R}/F\otimes_{\mathbb{Q}}\mathbb{R}}(b^*b)$ on each factor on the right, where it becomes $\mathrm{Tr}_{\mathbb{C}/\mathbb{R}}(\bar{z}z) = 2|z|^2 > 0$. Thus, we are in the PEL situation considered in §8.

Let E be a CM-field with largest real subfield F. Each embedding of F into $\mathbb{R}$ will extend to two conjugate embeddings of E into $\mathbb{C}$. A ***CM-type*** Φ for E is a choice of one element from each conjugate pair $\{\varphi, \overline{\varphi}\}$. In other words, it is a subset $\Phi \subset \mathrm{Hom}(E,\mathbb{C})$ such that

$$\mathrm{Hom}(E,\mathbb{C}) = \Phi \sqcup \overline{\Phi} \qquad (\text{disjoint union, } \overline{\Phi} = \{\overline{\varphi} \mid \varphi \in \Phi\}).$$

Because E is quadratic over F, $E = F[\alpha]$ with α a root of a polynomial $X^2 + aX + b$. On completing the square, we obtain an α such that $\alpha^2 \in F^\times$. Then $\alpha^* = -\alpha$. Such an element α of E is said to be ***totally imaginary*** (its image in $\mathbb{C}$ under every embedding is purely imaginary).

Abelian varieties of CM-type. Let E be a CM-field of degree $2g$ over $\mathbb{Q}$. Let A be an abelian variety of dimension g over $\mathbb{C}$, and let i be a homomorphism $E \to \mathrm{End}^0(A)$. If

$$\mathrm{Tr}(i(a) \mid \mathrm{Tgt}_0(A)) = \textstyle\sum_{\varphi \in \Phi} \varphi(a), \quad \text{all } a \in E, \tag{43}$$

for some CM-type Φ of E, then (A, i) is said to be of ***CM-type*** (E, Φ).

REMARK 10.1. (a) In fact, (A, i) will always be of CM-type for some Φ. Recall (p319) that $A(\mathbb{C}) \cong \mathrm{Tgt}_0(A)/\Lambda$ with Λ a lattice in $\mathrm{Tgt}_0(A)$ (so $\Lambda \otimes \mathbb{R} \cong \mathrm{Tgt}_0(A)$). Moreover,

$$\begin{aligned}
&\Lambda \otimes \mathbb{Q} \cong H_1(A, \mathbb{Q}) \\
&\Lambda \otimes \mathbb{R} \cong H_1(A, \mathbb{R}), \cong \mathrm{Tgt}_0(A) \\
&\Lambda \otimes \mathbb{C} = H_1(A, \mathbb{C}) \cong H^{-1,0} \oplus H^{0,-1} \cong \mathrm{Tgt}_0(A) \oplus \overline{\mathrm{Tgt}_0(A)}.
\end{aligned}$$

Now $H_1(A, \mathbb{Q})$ is a one-dimensional vector space over E, so $H_1(A, \mathbb{C}) \cong \bigoplus_{\varphi\colon E \to \mathbb{C}} \mathbb{C}_\varphi$ where $\mathbb{C}_\varphi$ denotes a 1-dimensional vector space with E acting through φ. If φ occurs in $\mathrm{Tgt}_0(A)$, then $\overline{\varphi}$ occurs in $\overline{\mathrm{Tgt}_0(A)}$, and so $\mathrm{Tgt}_0(A) \cong \bigoplus_{\varphi \in \Phi} \mathbb{C}_\varphi$ with Φ a CM-type for E.

(b) A field E of degree $2g$ over $\mathbb{Q}$ acting on a complex abelian variety A of dimension g need not be be CM unless A is simple.

Let Φ be a CM-type on E, and let $\mathbb{C}^\Phi$ be a direct sum of copies of $\mathbb{C}$ indexed by Φ. Denote by Φ again the homomorphism $\mathcal{O}_E \to \mathbb{C}^\Phi$, $a \mapsto (\varphi a)_{\varphi \in \Phi}$.

PROPOSITION 10.2. *The image $\Phi(\mathcal{O}_E)$ of $\mathcal{O}_E$ in $\mathbb{C}^\Phi$ is a lattice, and the quotient $\mathbb{C}^\Phi/\Phi(\mathcal{O}_E)$ is an abelian variety A_Φ of CM-type (E, Φ) for the natural homomorphism $i_\Phi\colon E \to \mathrm{End}^0(A_\Phi)$. Any other pair (A, i) of CM-type (E, Φ) is E-isogenous to (A_Φ, i_Φ).*

PROOF. We have

$$\mathcal{O}_E \otimes_\mathbb{Z} \mathbb{R} \cong \mathcal{O}_E \otimes_\mathbb{Z} \mathbb{Q} \otimes_\mathbb{Q} \mathbb{R} \cong E \otimes_\mathbb{Q} \mathbb{R} \xrightarrow[\cong]{e \otimes r \mapsto (\ldots, r \cdot \varphi e, \ldots)} \mathbb{C}^\Phi,$$

and so $\Phi(\mathcal{O}_E)$ is a lattice in $\mathbb{C}^\Phi$.

To show that the quotient is an abelian variety, we have to exhibit a riemann form (6.7). Let α be a totally imaginary element of E. The weak approximation theorem allows us to choose α so that $\Im(\varphi\alpha) > 0$ for $\varphi \in \Phi$, and we can multiply it by an integer (in $\mathbb{N}$) to make it an algebraic integer. Define

$$\psi(u, v) = \mathrm{Tr}_{E/\mathbb{Q}}(\alpha u v^*), \qquad u, v \in \mathcal{O}_E.$$

Then $\psi(u, v) \in \mathbb{Z}$. The remaining properties can be checked on the right of (42). Here ψ takes the form $\psi = \sum_{\varphi \in \Phi} \psi_\varphi$, where

$$\psi_\varphi(u, v) = \mathrm{Tr}_{\mathbb{C}/\mathbb{R}}(\alpha_\varphi \cdot u \cdot \overline{v}), \quad \alpha_\varphi = \varphi(\alpha), \quad u, v \in \mathbb{C}.$$

Because α is totally imaginary,

$$\psi_\varphi(u, v) = \alpha_\varphi(u\overline{v} - \overline{u}v) \in \mathbb{R},$$

from which it follows that $\psi_\varphi(u,u)=0$, $\psi_\varphi(iu,iv)=\psi_\varphi(u,v)$, and $\psi_\varphi(u,iu)>0$ for $u\neq 0$. Thus, ψ is a riemann form and A_Φ is an abelian variety.

An element $\alpha\in\mathcal{O}_E$ acts on $\mathbb{C}^\Phi$ as muliplication by $\Phi(\alpha)$. This preserves $\Phi(\mathcal{O}_E)$, and so defines a homomorphism $\mathcal{O}_E\to\mathrm{End}(A_\Phi)$. On tensoring this with $\mathbb{Q}$, we obtain the homomorphism i_Φ. The map $\mathbb{C}^\Phi\to\mathbb{C}^\Phi/\Phi(\mathcal{O}_E)$ defines an isomorphism $\mathbb{C}^\Phi=\mathrm{Tgt}_0(\mathbb{C}^\Phi)\to\mathrm{Tgt}_0(A_\Phi)$ compatible with the actions of E. Therefore, (A_Φ,i_Φ) is of CM-type (E,Φ).

Finally, let (A,i) be of CM-type (E,Φ). The condition (43) means that $\mathrm{Tgt}_0(A)$ is isomorphic to $\mathbb{C}^\Phi$ as an $E\otimes_\mathbb{Q}\mathbb{C}$-module. Therefore, $A(\mathbb{C})$ is a quotient of $\mathbb{C}^\Phi$ by a lattice Λ such that $\mathbb{Q}\Lambda$ is stable under the action of E on $\mathbb{C}^\Phi$ given by Φ (see 6.7 et seq.). This implies that $\mathbb{Q}\Lambda=\Phi(E)$, and so $\Lambda=\Phi(\Lambda')$ where Λ' is a lattice in E. Now, $N\Lambda'\subset\mathcal{O}_E$ for some N, and we have E-isogenies

$$\mathbb{C}^\Phi/\Lambda\xrightarrow{N}\mathbb{C}^\Phi/N\Lambda\leftarrow\mathbb{C}^\Phi/\Phi(\mathcal{O}_E).\qquad\square$$

PROPOSITION 10.3. *Let (A,i) be an abelian variety of CM-type (E,Φ) over $\mathbb{C}$. Then (A,i) has a model over $\mathbb{Q}^{al}$, uniquely determined up to isomorphism.*

PROOF. Let $k\subset\Omega$ be algebraically closed fields of characteristic zero. For an abelian variety A over k, the torsion points in $A(k)$ are zariski dense, and the map on torsion points $A(k)_{\mathrm{tors}}\to A(\Omega)_{\mathrm{tors}}$ is bijective (see (40)), and so every regular map $A_\Omega\to W_\Omega$ (W a variety over k) is fixed by the automorphisms of Ω/k and is therefore defined over k (AG 16.9; see also 13.1 below). It follows that $A\mapsto A_\Omega\colon\mathsf{AV}(k)\to\mathsf{AV}(\Omega)$ is fully faithful.

It remains to show that every abelian variety (A,i) of CM-type over $\mathbb{C}$ arises from a pair over $\mathbb{Q}^{\mathrm{al}}$. The polynomials defining A and i have coefficients in some subring R of $\mathbb{C}$ that is finitely generated over $\mathbb{Q}^{\mathrm{al}}$. According to the Hilbert Nullstellensatz, a maximal ideal $\mathfrak{m}$ of R will have residue field $\mathbb{Q}^{\mathrm{al}}$, and the reduction of (A,i) mod $\mathfrak{m}$ is called a ***specialization*** of (A,i). Any specialization (A',i') of (A,i) to a pair over $\mathbb{Q}^{\mathrm{al}}$ with A' nonsingular will still be of CM-type (E,Φ), and therefore (see 10.2) there exists an isogeny $(A',i')_\mathbb{C}\to(A,i)$. The kernel H of this isogeny is a subgroup of $A'(\mathbb{C})_{\mathrm{tors}}=A'(\mathbb{Q}^{\mathrm{al}})_{\mathrm{tors}}$, and $(A'/H,i)$ will be a model of (A,i) over $\mathbb{Q}^{\mathrm{al}}$. $\square$

REMARK 10.4. The proposition implies that, in order for an elliptic curve A over $\mathbb{C}$ to be of CM-type, its j-invariant must be algebraic.

Let A be an abelian variety of dimension g over a subfield k of $\mathbb{C}$, and let $i\colon E\to\mathrm{End}^0(A)$ be a homomorphism with E a CM-field of degree $2g$. Then $\mathrm{Tgt}_0(A)$ is a k-vector space of dimension g on which E acts k-linearly, and, provided k is large enough to contain all conjugates of E, it will decompose into one-dimensional k-subspaces indexed by a subset Φ of $\mathrm{Hom}(E,k)$. When we identify Φ with a subset of $\mathrm{Hom}(E,\mathbb{C})$, it becomes a CM-type, and we again say (A,i) is of ***CM-type*** (E,Φ).

Let A be an abelian variety over a number field K. We say that A has ***good reduction at*** $\mathfrak{P}$ if it extends to an abelian scheme over $\mathcal{O}_{K,\mathfrak{P}}$, i.e., a smooth proper scheme over $\mathcal{O}_{K,\mathfrak{P}}$ with a group structure. In down-to-earth terms this means the following: embed A as a closed subvariety of some projective space $\mathbb{P}^n_K$; for each polynomial $P(X_0,\ldots,X_n)$ in the homogeneous ideal $\mathfrak{a}$ defining $A\subset\mathbb{P}^n_K$, multiply P by an element of K so that it (just) lies in $\mathcal{O}_{K,\mathfrak{P}}[X_0,\ldots,X_n]$ and let $\overline{P}$ denote the reduction of P modulo $\mathfrak{P}$; the $\overline{P}$'s obtained in this fashion generate a homogeneous $\overline{\mathfrak{a}}$ ideal in $k[X_0,\ldots,X_n]$ where $k=\mathcal{O}_K/\mathfrak{P}$; the abelian variety A has good reduction at $\mathfrak{P}$ if it is possible to choose the projective embedding of A so that the zero set of $\overline{\mathfrak{a}}$ is an abelian variety $\overline{A}$ over k. Then $\overline{A}$ is called ***the reduction of** A **at** $\mathfrak{P}$*. It can be shown that, up to a canonical isomorphism, $\overline{A}$ is independent of all choices. For $\ell \neq \mathrm{char}(k)$, $V_\ell(A) \cong V_\ell(\overline{A})$. There is an injective homorphism $\mathrm{End}(A) \to \mathrm{End}(\overline{A})$ compatible with $V_\ell(A) \cong V_\ell(\overline{A})$ (both are reduction maps).

PROPOSITION 10.5. *Let (A, i) be an abelian variety of CM-type (E, Φ) over a number field $K \subset \mathbb{C}$, and let $\mathfrak{P}$ be a prime ideal in $\mathcal{O}_K$. Then, after possibly replacing K by a finite extension, A will have good reduction at $\mathfrak{P}$.*

PROOF. We use the Néron (alias, Ogg-Shafarevich) criterion (Serre and Tate 1968, Theorem 1):

> an abelian variety over a number field K has good reduction at $\mathfrak{P}$ if for some prime $\ell \neq \mathrm{char}(\mathcal{O}_K/\mathfrak{P})$, the inertia group I at $\mathfrak{P}$ acts trivially on $T_\ell A$.

In our case, $V_\ell A$ is a free $E \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$-module of rank 1 because $H_1(A_{\mathbb{C}}, \mathbb{Q})$ is a one-dimensional vector space over E and $V_\ell A \cong H_1(A_{\mathbb{C}}, \mathbb{Q}) \otimes \mathbb{Q}_\ell$ (see (41)). Therefore, $E \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ is its own centralizer in $\mathrm{End}_{\mathbb{Q}_\ell}(V_\ell A)$ and the representation of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ on $V_\ell A$ has image in $(E \otimes \mathbb{Q}_\ell)^\times$, and, in fact, in a compact subgroup of $(E \otimes \mathbb{Q}_\ell)^\times$. But such a subgroup will have a pro-ℓ subgroup of finite index. Since I has a pro-p subgroup of finite index (ANT, 7.5), this shows that image of I is finite. After K has been replaced by a finite extension, the image of I will be trivial, and Néron's criterion applies. □

Abelian varieties over a finite field. Let $\mathbb{F}$ be an algebraic closure of the field $\mathbb{F}_p$ of p-elements, and let $\mathbb{F}_q$ be the subfield of $\mathbb{F}$ with $q = p^m$ elements. An element a of $\mathbb{F}$ lies in $\mathbb{F}_q$ if and only if $a^q = a$. Recall that, in characteristic p, $(X+Y)^p = X^p + Y^p$. Therefore, if $f(X_1, \ldots, X_n)$ has coefficients in $\mathbb{F}_q$, then

$$f(X_1, \ldots, X_n)^q = f(X_1^q, \ldots, X_n^q), \quad f(a_1, \ldots, a_n)^q = f(a_1^q, \ldots, a_n^q), \quad a_i \in \mathbb{F}.$$

In particular,

$$f(a_1, \ldots, a_n) = 0 \implies f(a_1^q, \ldots, a_n^q) = 0, \quad a_i \in \mathbb{F}.$$

PROPOSITION 10.6. *There is a unique way to attach to every variety V over $\mathbb{F}_q$ a regular map $\pi_V : V \to V$ such that*

(a) *for any regular map $\alpha : V \to W$, $\alpha \circ \pi_V = \pi_W \circ \alpha$;*
(b) *$\pi_{\mathbb{A}^n}$ is the map $(a_1, \ldots, a_n) \mapsto (a_1^q, \ldots, a_n^q)$.*

PROOF. For an affine variety $V = \mathrm{Specm} A$, define π_V be the map corresponding to the $\mathbb{F}_q$-homomorphism $x \mapsto x^q : A \to A$. The rest of the proof is straightforward. □

The map π_V is called the ***Frobenius map of** V*.

THEOREM 10.7 (Weil 1948). *For an abelian variety A over $\mathbb{F}_q$, $\mathrm{End}^0(A)$ is a finite-dimensional semisimple $\mathbb{Q}$-algebra with π_A in its centre. For every embedding $\rho : \mathbb{Q}[\pi_A] \to \mathbb{C}$, $|\rho(\pi_A)| = q^{\frac{1}{2}}$.*

PROOF. See, for example, Milne 1986, 19.1. □

If A is simple, $\mathbb{Q}[\pi_A]$ is a field (p334), and π_A is an algebraic integer in it (p334). An algebraic integer π such that $|\rho(\pi)| = q^{\frac{1}{2}}$ for all embeddings $\rho : \mathbb{Q}[\pi] \to \mathbb{C}$ is called a ***Weil** q**-integer*** (formerly, Weil q-number).

For a Weil q-integer π,

$$\rho(\pi)\cdot\overline{\rho(\pi)} = q = \rho(\pi)\cdot\rho(q/\pi), \quad \text{all } \rho\colon \mathbb{Q}[\pi]\to\mathbb{C},$$

and so $\rho(q/\pi) = \overline{\rho(\pi)}$. It follows that the field $\rho(\mathbb{Q}[\pi])$ is stable under complex conjugation and that the automorphism of $\mathbb{Q}[\pi]$ induced by complex conjugation sends π to q/π and is independent of ρ. This implies that $\mathbb{Q}[\pi]$ is a CM-field (the typical case), $\mathbb{Q}$, or $\mathbb{Q}[\sqrt{p}]$.

LEMMA 10.8. *Let π and π' be Weil q-integers lying in the same field E. If $\mathrm{ord}_v(\pi) = \mathrm{ord}_v(\pi')$ for all $v|p$, then $\pi' = \zeta\pi$ for some root of 1 in E.*

PROOF. As noted above, there is an automorphism of $\mathbb{Q}[\pi]$ sending π to q/π. Therefore q/π is also an algebraic integer, and so $\mathrm{ord}_v(\pi) = 0$ for every finite $v \nmid p$. Since the same is true for π', we find that $|\pi|_v = |\pi'|_v$ for all v. Hence π/π' is a unit in $\mathcal{O}_E$ such that $|\pi/\pi'|_v = 1$ for all $v|\infty$. But in the course of proving the unit theorem, one shows that such a unit has to be root of 1 (ANT, 5.6). □

The Shimura-Taniyama formula.

LEMMA 10.9. *Let (A, i) be an abelian variety of CM-type (E, Φ) over a number field $k \subset \mathbb{C}$ having good reduction at $\mathfrak{P} \subset \mathcal{O}_k$ to $(\overline{A}, \bar{\imath})$ over $\mathcal{O}_k/\mathfrak{P} = \mathbb{F}_q$. Then the Frobenius map $\pi_{\overline{A}}$ of $\overline{A}$ lies in $\bar{\imath}(E)$.*

PROOF. Let $\pi = \pi_{\overline{A}}$. It suffices to check that π lies in $\bar{\imath}(E)$ after tensoring with $\mathbb{Q}_\ell$. As we saw in the proof of (10.5), $V_\ell A$ is a free $E\otimes_{\mathbb{Q}}\mathbb{Q}_\ell$-module of rank 1. It follows that $V_\ell\overline{A}$ is also a free $E\otimes_{\mathbb{Q}}\mathbb{Q}_\ell$-module of rank 1 (via $\bar{\imath}$). Therefore, any endomorphism of $V_\ell\overline{A}$ commuting with the action of $E\otimes\mathbb{Q}_\ell$ will lie in $E\otimes\mathbb{Q}_\ell$. □

Thus, from (A, i) and a prime $\mathfrak{P}$ of k at which A has good reduction, we get a Weil q-integer $\pi \in E$.

THEOREM 10.10 (Shimura-Taniyama). *In the situation of the lemma, assume that k is galois over $\mathbb{Q}$ and contains all conjugates of E. Then for all primes v of E dividing p,*

$$\frac{\mathrm{ord}_v(\pi)}{\mathrm{ord}_v(q)} = \frac{|\Phi\cap H_v|}{|H_v|} \tag{44}$$

where $H_v = \{\rho\colon E\to k \mid \rho^{-1}(\mathfrak{P}) = \mathfrak{p}_v\}$ and $|S|$ denotes the order of a set S.

REMARK 10.11. (a) According to (10.8), the theorem determines π up to a root of 1. Note that the formula depends only on (E, Φ). It is possible to see directly that different pairs (A, i) over k of CM-type (E, Φ) can give different Frobenius elements, but they will differ only by a root of 1.

(b) Let $*$ denote complex conjugation on $\mathbb{Q}[\pi]$. Then $\pi\pi^* = q$, and so

$$\mathrm{ord}_v(\pi) + \mathrm{ord}_v(\pi^*) = \mathrm{ord}_v(q). \tag{45}$$

Moreover,

$$\mathrm{ord}_v(\pi^*) = \mathrm{ord}_{v^*}(\pi)$$

and

$$\Phi\cap H_{v^*} = \overline{\Phi}\cap H_v.$$

Therefore, (44) is consistent with (45):

$$\frac{\mathrm{ord}_v(\pi)}{\mathrm{ord}_v(q)} + \frac{\mathrm{ord}_v(\pi^*)}{\mathrm{ord}_v(q)} \overset{(44)}{=} \frac{|\Phi\cap H_v| + |\Phi\cap H_{v^*}|}{|H_v|} = \frac{|(\Phi\cup\overline{\Phi})\cap H_v|}{|H_v|} = 1.$$

In fact, (44) is the only obvious formula for $\mathrm{ord}_v(\pi)$ consistent with (45), which is probably a more convincing argument for its validity than the proof sketched below.

The $\mathcal{O}_E$-structure of the tangent space. Let R be a Dedekind domain. Any finitely generated torsion R-module M can be written as a direct sum $\bigoplus_i R/\mathfrak{p}_i^{r_i}$ with each $\mathfrak{p}_i$ an ideal in R, and the set of pairs $(\mathfrak{p}_i, r_i)$ is uniquely determined by M. Define $|M|_R = \prod \mathfrak{p}_i^{r_i}$. For example, for $R = \mathbb{Z}$, M is a finite abelian group and $|M|_{\mathbb{Z}}$ is the ideal in $\mathbb{Z}$ generated by the order of M.

For Dedekind domains $R \subset S$ with S finite over R, there is a norm homomorphism sending fractional ideals of S to fractional ideals of R (ANT, p58). It is compatible with norms of elements, and

$$\mathrm{Nm}(\mathfrak{P}) = \mathfrak{p}^{f(\mathfrak{P}/\mathfrak{p})}, \quad \mathfrak{P} \text{ prime}, \ \mathfrak{p} = \mathfrak{P} \cap R$$

Clearly,

$$|S/\mathfrak{A}|_R = \mathrm{Nm}(\mathfrak{A}) \tag{46}$$

since this is true for prime ideals, and both sides are multiplicative.

PROPOSITION 10.12. *Let A be an abelian variety of dimension g over $\mathbb{F}_q$, and let i be a homomorphism from the ring of integers $\mathcal{O}_E$ of a field E of degree $2g$ over $\mathbb{Q}$ into $\mathrm{End}(A)$. Then*

$$|\operatorname{Tgt}_0 A|_{\mathcal{O}_E} = (\pi_A).$$

PROOF. Omitted (for a scheme-theoretic proof, see Giraud 1968, Théorème 1). □

Sketch of the proof the Shimura-Taniyama formula. We return to the situation of the Theorem 10.10. After replacing A with an isogenous variety, we may assume $i(\mathcal{O}_E) \subset \mathrm{End}(A)$. By assumption, there exists an abelian scheme $\mathcal{A}$ over $\mathcal{O}_{k,\mathfrak{P}}$ with generic fibre A and special fibre an abelian variety $\overline{A}$. Because $\mathcal{A}$ is smooth over $\mathcal{O}_{k,\mathfrak{P}}$, the relative tangent space of $\mathcal{A}/\mathcal{O}_{k,\mathfrak{P}}$ is a free $\mathcal{O}_{k,\mathfrak{P}}$-module T of rank g endowed with an action of $\mathcal{O}_E$ such that

$$T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} k = \operatorname{Tgt}_0(A), \quad T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} \mathcal{O}_{k,\mathfrak{P}}/\mathfrak{P} = \operatorname{Tgt}_0(\overline{A}).$$

Therefore,

$$(\pi) \overset{10.12}{=} \left|\operatorname{Tgt}_0 \overline{A}\right|_{\mathcal{O}_E} = \left|T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} (\mathcal{O}_{k,\mathfrak{P}}/\mathfrak{P})\right|_{\mathcal{O}_E}. \tag{47}$$

For simplicity, assume that $(p) =_{\mathrm{df}} \mathfrak{P} \cap \mathbb{Z}$ is unramified in E. Then the isomorphism of E-modules

$$T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} k \approx k^{\Phi}$$

induces an isomorphism of $\mathcal{O}_E$-modules

$$T \approx \mathcal{O}_{k,\mathfrak{P}}^{\Phi} \tag{48}$$

In other words, T is a direct sum of copies of $\mathcal{O}_{k,\mathfrak{P}}$ indexed by the elements of Φ, and $\mathcal{O}_E$ acts on the φ^{th} copy through the map

$$\mathcal{O}_E \xrightarrow{\varphi} \mathcal{O}_k \subset \mathcal{O}_{k,\mathfrak{P}}.$$

As $\mathcal{O}_k/\mathfrak{P} \cong \mathcal{O}_{k,\mathfrak{P}}/\mathfrak{P}$ (ANT, 3.11), the contribution of the φ^{th} copy to (π) in (47) is

$$|\mathcal{O}_k/\mathfrak{P}|_{\mathcal{O}_E} \overset{(46)}{=} \varphi^{-1}(\mathrm{Nm}_{k/\varphi E}\,\mathfrak{P}).$$

Thus,

$$(\pi) = \textstyle\prod_{\varphi\in\Phi}\varphi^{-1}(\mathrm{Nm}_{k/\varphi E}\,\mathfrak{P}). \tag{49}$$

It is only an exercise to derive (44) from (49).

NOTES. The original formulation of the Shimura-Taniyama theorem is in fact (49). It is proved in Shimura and Taniyama 1961, III.13, in the unramified case using spaces of differentials rather than tangent spaces. The proof sketched above is given in detail in Giraud 1968, and there is a proof using p-divisible groups in Tate 1969, §5. See also Serre 1968, pII-28.

11. Complex multiplication: the main theorem

Review of class field theory. Classical class field theory classifies the abelian extensions of a number field E, i.e., the galois extensions L/E such $\mathrm{Gal}(L/E)$ is commutative. Let E^{ab} be the composite of all the finite abelian extensions of E inside some fixed algebraic closure E^{al} of E. Then E^{ab} is an infinite galois extension of E.

According to class field theory, there exists a continuous surjective homomorphism (the ***reciprocity*** or ***Artin map***)

$$\mathrm{rec}_E\colon \mathbb{A}_E^\times \to \mathrm{Gal}(E^{\mathrm{ab}}/E)$$

such that, for every finite extension L of E contained in E^{ab}, rec_E gives rise to a commutative diagram

$$\begin{array}{ccc}
E^\times\backslash\mathbb{A}_E^\times & \xrightarrow[\text{onto}]{\mathrm{rec}_E} & \mathrm{Gal}(E^{\mathrm{ab}}/E) \\
\downarrow & & \downarrow{\scriptstyle \sigma\mapsto\sigma|L} \\
E^\times\backslash\mathbb{A}_E^\times/\mathrm{Nm}_{L/E}(\mathbb{A}_L^\times) & \xrightarrow[\cong]{\mathrm{rec}_{L/E}} & \mathrm{Gal}(L/E).
\end{array}$$

It is determined by the following two properties:

(a) $\mathrm{rec}_{L/E}(u) = 1$ for every $u = (u_v) \in \mathbb{A}_E^\times$ such that
  i) if v is unramified in L, then u_v is a unit,
  ii) if v is ramified in L, then u_v is sufficiently close to 1 (depending only on L/E), and
  iii) if v is real but becomes complex in L, then $u_v > 0$.

(b) For every prime v of E unramified in L, the idèle

$$\alpha = (1,\ldots,1,\underset{v}{\pi},1,\ldots), \quad \pi \text{ a prime element of } \mathcal{O}_{E_v},$$

maps to the Frobenius element $(v, L/E) \in \mathrm{Gal}(L/E)$.

Recall that if $\mathfrak{P}$ is a prime ideal of L lying over $\mathfrak{p}_v$, then $(v, L/E)$ is the automorphism of L/E fixing $\mathfrak{P}$ and acting as $x \mapsto x^{(\mathcal{O}_E\,:\,\mathfrak{p}_v)}$ on $\mathcal{O}_L/\mathfrak{P}$.

To see that there is at most one map satisfying these conditions, let $\alpha \in \mathbb{A}_E^\times$, and use the weak approximation theorem to choose an $a \in E^\times$ that is close to α_v for all primes v that ramify in L or become complex. Then $\alpha = au\beta$ with u an idèle as in (a) and β a finite product of idèles as in (b). Now $\mathrm{rec}_{L/E}(\alpha) = \mathrm{rec}_{L/E}(\beta)$, which can be computed using (b).

Note that, because $\mathrm{Gal}(E^{\mathrm{ab}}/E)$ is totally disconnected, the identity component of $E^\times\backslash\mathbb{A}_E^\times$ is contained in the kernel of rec_E. In particular, the identity component of $\prod_{v|\infty} E_v^\times$ is contained in the kernel, and so, when E is totally imaginary, rec_E factors through $E^\times\backslash\mathbb{A}_{E,f}^\times$.

For $E = \mathbb{Q}$, the reciprocity map factors through $\mathbb{Q}^\times\backslash\{\pm\} \times \mathbb{A}_f^\times$, and every element in this quotient is uniquely represented by an element of $\hat{\mathbb{Z}}^\times \subset \mathbb{A}_f^\times$. In this case, we get the diagram

$$\begin{array}{ccccc} \hat{\mathbb{Z}}^\times & \xrightarrow[\cong]{\mathrm{rec}_\mathbb{Q}} & \mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}) & = & \bigcup \mathbb{Q}[\zeta_N] \\ \downarrow & & \downarrow{\scriptstyle \mathrm{restrict}} & & \\ (\mathbb{Z}/N\mathbb{Z})^\times & \xrightarrow[\cong]{[a]\mapsto(\zeta_N\mapsto\zeta_N^a)} & \mathrm{Gal}(\mathbb{Q}[\zeta_N]/\mathbb{Q}) & & \end{array} \tag{50}$$

which *commutes with an inverse.* This can be checked by writing an idèle α in the form $au\beta$ as above, but it is more instructive to look at an example. Let p be a prime not dividing N, and let

$$\alpha = p\cdot(1,\ldots,1,\underset{p}{p^{-1}},1,\ldots) \in \mathbb{Z}\cdot\mathbb{A}_f^\times = \mathbb{A}_f^\times.$$

Then $\alpha \in \hat{\mathbb{Z}}^\times$ and has image $[p]$ in $\mathbb{Z}/N\mathbb{Z}$, which acts as $(p, \mathbb{Q}[\zeta_N]/\mathbb{Q})$ on $\mathbb{Q}[\zeta_N]$. On the other hand, $\mathrm{rec}_\mathbb{Q}(\alpha) = \mathrm{rec}_\mathbb{Q}((1,\ldots,p^{-1},\ldots))$, which acts as $(p,\mathbb{Q}[\zeta_N]/\mathbb{Q})^{-1}$.

NOTES. For the proofs of the above statements, see Tate 1967 or my notes CFT.

Convention for the (Artin) reciprocity map. It simplifies the formulas in Langlands theory if one replaces the reciprocity map with its reciprocal. For $\alpha \in \mathbb{A}_E^\times$, write

$$\mathrm{art}_E(\alpha) = \mathrm{rec}_E(\alpha)^{-1}. \tag{51}$$

Now, the diagram (50) commutes. In other words,

$$\mathrm{art}_\mathbb{Q}(\chi(\sigma)) = \sigma, \quad \text{for } \sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}),$$

where χ is the cyclotomic character $\mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}) \to \hat{\mathbb{Z}}^\times$, which is characterized by

$$\sigma\zeta = \zeta^{\chi(\sigma)}, \quad \zeta \text{ a root of 1 in } \mathbb{C}^\times.$$

The reflex field and norm of a CM-type. Let (E,Φ) be a CM-type.

DEFINITION 11.1. The ***reflex field*** E^* of (E,Φ) is the subfield of $\mathbb{Q}^{\mathrm{al}}$ characterized by any one of the following equivalent conditions:

(a) $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ fixes E^* if and only if $\sigma\Phi = \Phi$; here $\sigma\Phi = \{\sigma\circ\varphi | \varphi\in\Phi\}$;
(b) E^* is the field generated over $\mathbb{Q}$ by the elements $\sum_{\varphi\in\Phi}\varphi(a)$, $a \in E$;
(c) E^* is the smallest subfield k of $\mathbb{Q}^{\mathrm{al}}$ such that there exists a k-vector space V with an action of E for which

$$\mathrm{Tr}_k(a|V) = \textstyle\sum_{\varphi\in\Phi}\varphi(a), \quad \text{all } a \in E.$$

Let V be an E^*-vector space with an action of E such that $\mathrm{Tr}_{E^*}(a|V) = \sum_{\varphi\in\Phi}\varphi(a)$ for all $a \in E$. We can regard V as an $E^*\otimes_\mathbb{Q} E$-space, or as an E-vector space with a E-linear action of E^*. The ***reflex norm*** is the homomorphism $N_{\Phi^*}\colon (\mathbb{G}_m)_{E^*/\mathbb{Q}} \to (\mathbb{G}_m)_{E/\mathbb{Q}}$ such that

$$N_{\Phi^*}(a) = \det{}_E(a|V), \quad \text{all } a \in E^{*\times}.$$

This definition is independent of the choice of V because V is unique up to an isomorphism respecting the actions of E and E^*.

Let (A,i) be an abelian variety of CM-type (E,Φ) defined over $\mathbb{C}$. According to (11.1c) applied to $\mathrm{Tgt}_0(A)$, any field of definition of (A,i) contains E^*.

Statement of the main theorem of complex multiplication. A homomorphism $\sigma\colon k \to \Omega$ of fields defines a functor $V \mapsto \sigma V$, $\alpha \mapsto \sigma\alpha$, "extension of the base field" from varieties over k to varieties over Ω. In particular, an abelian variety A over k equipped with a homomorphism $i\colon E \to \mathrm{End}^0(A)$ defines a similar pair $\sigma(A,i) = (\sigma A, {}^\sigma i)$ over Ω. Here ${}^\sigma i\colon E \to \mathrm{End}(\sigma A)$ is defined by

$$ {}^\sigma i(a) = \sigma(i(a)). $$

A point $P \in A(k)$ gives a point $\sigma P \in A(\Omega)$, and so σ defines a homomorphism $\sigma\colon V_f(A) \to V_f(\sigma A)$ provided that k and Ω are algebraically closed (otherwise one would have to choose an extension of k to a homomorphism $k^{\mathrm{al}} \to \Omega^{\mathrm{al}}$).

THEOREM 11.2. *Let (A,i) be an abelian variety of CM-type (E,Φ) over $\mathbb{C}$, and let $\sigma \in \mathrm{Aut}(\mathbb{C}/E^*)$. For any $s \in \mathbb{A}^\times_{E^*,f}$ with $\mathrm{art}_{E^*}(s) = \sigma|E^{*ab}$, there is a unique E-linear isogeny $\alpha\colon A \to \sigma A$ such that $\alpha(N_{\Phi^*}(s)\cdot x) = \sigma x$ for all $x \in V_f A$.*

PROOF. Formation of the tangent space commutes with extension of the base field, and so

$$ \mathrm{Tgt}_0(\sigma A) = \mathrm{Tgt}_0(A) \otimes_{\mathbb{C},\sigma} \mathbb{C} $$

as an $E \otimes_{\mathbb{Q}} \mathbb{C}$-module. Therefore, $(\sigma A, {}^\sigma i)$ is of CM type $\sigma\Phi$. Since σ fixes E^*, $\sigma\Phi = \Phi$, and so there exists an E-linear isogeny $\alpha\colon A \to \sigma A$ (10.2). The map

$$ V_f(A) \xrightarrow{\sigma} V_f(\sigma A) \xrightarrow{V_f(\alpha)^{-1}} V_f(A) $$

is $E\otimes_{\mathbb{Q}}\mathbb{A}_f$-linear. As $V_f(A)$ is free of rank one over $E\otimes_{\mathbb{Q}}\mathbb{A}_f = \mathbb{A}_{E,f}$, this map must be multiplication by an element of $a \in \mathbb{A}^\times_{E,f}$. When the choice of α is changed, then a is changed only by an element of $E^\times$, and so we have a well-defined map

$$ \sigma \mapsto aE^\times\colon\ \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*) \to \mathbb{A}^\times_{E,f}/E^\times, $$

which one checks to be a homomorphism. The map factors through $\mathrm{Gal}(E^{*\mathrm{ab}}/E^*)$, and so, when composed with the reciprocity map art_{E^*}, it gives a homomorphism

$$ \eta\colon \mathbb{A}^\times_{E^*,f}/E^{*\times} \to \mathbb{A}^\times_{E,f}/E^\times. $$

We have to check that η is the homomorphism defined by N_{Φ^*}, but it can be shown that this follows from the Shimura-Taniyama formula (Theorem 10.10). The uniqueness follows from the faithfulness of the functor $A \mapsto V_f(A)$. □

REMARK 11.3. (a) If s is replaced by as, $a \in E^{*\times}$, then α must be replaced by $\alpha \circ N_{\Phi^*}(a)^{-1}$.

(b) The theorem is a statement about the E-isogeny class of (A,i). If $\beta\colon (A,i) \to (B,j)$ is an E-linear isogeny, and α satisfies the conditions of the theorem for (A,i), then $(\sigma\beta)\circ\alpha\circ\beta^{-1}$ satisfies the conditions for (B,j).

ASIDE 11.4. What happens in (11.2) when σ is not assumed to fix E^*? This also is known, thanks to Deligne and Langlands. For a discussion of this, and much else concerning complex multiplication, see my notes Milne 1979.

12. Definition of canonical models

We attach to each Shimura datum (G, X) an algebraic number field $E(G, X)$, and we define the canonical model of $\mathrm{Sh}(G, X)$ to be an inverse system of varieties over $E(G, X)$ that is characterized by reciprocity laws at certain special points.

Models of varieties. Let k be a subfield of a field Ω, and let V be a variety over Ω. A ***model*** of V over k (or a k-***structure*** on V) is a variety V_0 over k together with an isomorphism $\varphi\colon V_{0\Omega} \to V$. We often omit the map φ and regard a model as a variety V_0 over k such that $V_{0\Omega} = V$.

Consider an affine variety V over $\mathbb{C}$ and a subfield k of $\mathbb{C}$. An embedding $V \hookrightarrow \mathbb{A}^n_{\mathbb{C}}$ defines a model of V over k if the ideal $I(V)$ of polynomials zero on V is generated by polynomials in $k[X_1, \ldots, X_n]$, because then $I_0 =_{\mathrm{df}} I(V) \cap k[X_1, \ldots, X_n]$ is a radical ideal, $k[X_1, \ldots, X_n]/I_0$ is an affine k-algebra, and $V(I_0) \subset \mathbb{A}^n_k$ is a model of V. Moreover, every model (V_0, φ) arises in this way because every model of an affine variety is affine. However, different embeddings in affine space will usually give rise to different models. For example, the embeddings

$$\mathbb{A}^2_{\mathbb{C}} \xleftarrow{(x,y) \longleftarrow\!\shortmid\, (x,y)} V(X^2+Y^2-1) \xrightarrow{(x,y) \longmapsto (x,y/\sqrt{2})} \mathbb{A}^2_{\mathbb{C}}$$

define the $\mathbb{Q}$-structures

$$X^2 + Y^2 = 1, \quad X^2 + 2Y^2 = 1$$

on the curve $X^2 + Y^2 = 1$. These are not isomorphic.

Similar remarks apply to projective varieties.

In general, a variety over $\mathbb{C}$ will not have a model over a number field, and when it does, it will have many. For example, an elliptic curve E over $\mathbb{C}$ has a model over a number field if and only if its j-invariant $j(E)$ is an algebraic number, and if $Y^2Z = X^3 + aXZ^2 + bZ^3$ is one model of E over a number field k (meaning, $a, b \in k$), then $Y^2Z = X^3 + ac^2XZ^2 + bc^3Z^3$ is a second, which is isomorphic to the first only if c is a square in k.

The reflex field. For a reductive group G over $\mathbb{Q}$ and a subfield k of $\mathbb{C}$, we write $\mathcal{C}(k)$ for the set of $G(k)$-conjugacy classes of cocharacters of G_k defined over k:

$$\mathcal{C}(k) = G(k)\backslash \operatorname{Hom}(\mathbb{G}_m, G_k).$$

A homomorphism $k \to k'$ induces a map $\mathcal{C}(k) \to \mathcal{C}(k')$; in particular, $\operatorname{Aut}(k'/k)$ acts on $\mathcal{C}(k')$.

LEMMA 12.1. *Assume G splits over k, so that it contains a split maximal torus T, and let W be the Weyl group $N_{G(k)}(T)/C_{G(k)}(T)$ of T. Then the map*

$$W\backslash \operatorname{Hom}(\mathbb{G}_m, T_k) \to G(k)\backslash \operatorname{Hom}(\mathbb{G}_m, G_k)$$

is bijective.

PROOF. As any two maximal split tori are conjugate (Springer 1998, 15.2.6), the map is surjective. Let μ and μ' be cocharacters of T that are conjugate by an element of $G(k)$, say, $\mu = \operatorname{ad}(g) \circ \mu'$ with $g \in G(k)$. Then $\operatorname{ad}(g)(T)$ and T are both maximal split tori in the centralizer C of $\mu(\mathbb{G}_m)$, which is a connected reductive group (ibid., 15.3.2). Therefore, there exists a $c \in C(k)$ such that $\operatorname{ad}(cg)(T) = T$. Now cg normalizes T and $\operatorname{ad}(cg) \circ \mu' = \mu$, which proves that μ and μ' are in the same W-orbit. $\square$

Let (G,X) be a Shimura datum. For each $x\in X$, we have a cocharacter μ_x of $G_{\mathbb{C}}$:

$$\mu_x(z)=h_{x\mathbb{C}}(z,1).$$

A different $x\in X$ will give a conjugate μ_x, and so X defines an element $c(X)$ of $\mathcal{C}(\mathbb{C})$. Neither $\mathrm{Hom}(\mathbb{G}_m,T_{\mathbb{Q}^{\mathrm{al}}})$ nor W changes when we replace $\mathbb{C}$ with the algebraic closure $\mathbb{Q}^{\mathrm{al}}$ of $\mathbb{Q}$ in $\mathbb{C}$, and so the lemma shows that $c(X)$ contains a μ defined over $\mathbb{Q}^{\mathrm{al}}$ and that the $G(\mathbb{Q}^{\mathrm{al}})$-conjugacy class of μ is independent of the choice of μ. This allows us to regard $c(X)$ as an element of $\mathcal{C}(\mathbb{Q}^{\mathrm{al}})$.

DEFINITION 12.2. The ***reflex (*****or** ***dual) field*** $E(G,X)$ is the field of definition of $c(X)$, i.e., it is the fixed field of the subgroup of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ fixing $c(X)$ as an element of $\mathcal{C}(\mathbb{Q}^{\mathrm{al}})$ (or stabilizing $c(X)$ as a subset of $\mathrm{Hom}(\mathbb{G}_m,G_{\mathbb{Q}^{\mathrm{al}}})$).

Note that the reflex field a subfield of $\mathbb{C}$.

REMARK 12.3. (a) Any subfield k of $\mathbb{Q}^{\mathrm{al}}$ splitting G contains $E(G,X)$. This follows from the lemma, because $W\backslash\mathrm{Hom}(\mathbb{G}_m,T)$ does not change when we pass from k to $\mathbb{Q}^{\mathrm{al}}$. If follows that $E(G,X)$ has finite degree over $\mathbb{Q}$.

(b) If $c(X)$ contains a μ defined over k, then $k\supset E(G,X)$. Conversely, if G is quasi-split over k and $k\supset E(G,X)$, then $c(X)$ contains a μ defined over k (Kottwitz 1984, 1.1.3).

(c) Let $(G,X)\overset{i}{\hookrightarrow}(G',X')$ be an inclusion of Shimura data. Suppose σ fixes $c(X)$, and let $\mu\in c(X)$. Then $\sigma\mu=g\cdot\mu\cdot g^{-1}$ for some $g\in G(\mathbb{Q}^{\mathrm{al}})$, and so, for any $g'\in G'(\mathbb{Q}^{\mathrm{al}})$,

$$\sigma(g'\cdot(i\circ\mu)\cdot g'^{-1})=(\sigma g')(i(g))\cdot i\circ\mu\cdot(i(g))^{-1}(\sigma g')^{-1}\in c(X').$$

Hence σ fixes $c(X')$, and we have shown that

$$E(G,X)\supset E(G',X').$$

EXAMPLE 12.4. (a) Let T be a torus over $\mathbb{Q}$, and let h be a homomorphism $\mathbb{S}\to T_{\mathbb{R}}$. Then $E(T,h)$ is the field of definition of μ_h, i.e., the smallest subfield of $\mathbb{C}$ over which μ_h is defined.

(b) Let (E,Φ) be a CM-type, and let T be the torus $(\mathbb{G}_m)_{E/\mathbb{Q}}$, so that $T(\mathbb{Q})=E^\times$ and

$$T(\mathbb{R})=(E\otimes_{\mathbb{Q}}\mathbb{R})^\times\cong(\mathbb{C}^\Phi)^\times,\quad (e\otimes r)\mapsto(\varphi(e)\cdot r)_{\varphi\in\Phi}.$$

Define $h_\Phi\colon\mathbb{C}^\times\to T(\mathbb{R})$ to be $z\mapsto(z,\dots,z)$. The corresponding cocharacter μ_Φ is

$$\begin{array}{ccccc}\mathbb{C}^\times & \to & T(\mathbb{C}) & \cong & (\mathbb{C}^\Phi)^\times\times(\mathbb{C}^{\overline{\Phi}})^\times\\ z & \mapsto & & & (z,\dots,z,1,\dots,1)\end{array}$$

Therefore, $\sigma\mu_\Phi=\mu_\Phi$ if and only if σ stabilizes Φ, and so $E(T,h_\Phi)$ is the reflex field of (E,Φ) defined in (11.1).

(c) If (G,X) is a simple PEL datum of type (A) or (C), then $E(G,X)$ is the field generated over $\mathbb{Q}$ by $\{\mathrm{Tr}_X(b)\mid b\in B\}$ (Deligne 1971*c*, 6.1).

(d) Let (G,X) be the Shimura datum attached to a quaternion algebra B over a totally real number field F, as in Example 5.24. Then $c(X)$ is represented by the cocharacter μ:

$$\begin{array}{ccccc}G(\mathbb{C}) & \approx & \mathrm{GL}_2(\mathbb{C})^{I_{\mathrm{c}}} & \times & \mathrm{GL}_2(\mathbb{C})^{I_{\mathrm{nc}}}\\ \mu(z) & = & (1,\dots,1) & \times & \left(\left(\begin{smallmatrix}z&0\\0&1\end{smallmatrix}\right),\dots,\left(\begin{smallmatrix}z&0\\0&1\end{smallmatrix}\right)\right).\end{array}$$

Therefore, $E(G,X)$ is the fixed field of the stabilizer in $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ of $I_{\mathrm{nc}}\subset I$. For example, if I_{nc} consists of a single element v (so we have a Shimura curve), then $E(G,X)=v(F)$.

(e) When G is adjoint, $E(G,X)$ can be described as follows. Choose a maximal torus T in $G_{\mathbb{Q}^{\mathrm{al}}}$ and a base $(\alpha_i)_{i\in I}$ for the roots. Recall that the nodes of the dynkin diagram Δ of (G,T) are indexed by I. The galois group $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ acts on Δ. Each $c\in\mathcal{C}(\mathbb{Q}^{\mathrm{al}})$ contains a $\mu\colon\mathbb{G}_m\to G_{\mathbb{Q}^{\mathrm{al}}}$ such that $\langle\alpha_i,\mu\rangle\geq 0$ for all i (cf. 1.25), and the map

$$c\mapsto(\langle\alpha_i,\mu\rangle)_{i\in I}\colon\mathcal{C}(\mathbb{Q}^{\mathrm{al}})\to\mathbb{N}^I\qquad\text{(copies of }\mathbb{N}\text{ indexed by }I)$$

is a bijection. Therefore, $E(G,X)$ is the fixed field of the subgroup of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ fixing $(\langle\alpha_i,\mu\rangle)_{i\in I}\in\mathbb{N}^I$. It is either totally real or CM (Deligne 1971*b*, p139).

(f) Let (G,X) be a Shimura datum, and let $G\xrightarrow{\nu}T$ be the quotient of G by G^{der}. From (G,X), we get Shimura data $(G^{\mathrm{ad}},X^{\mathrm{ad}})$ and (T,h) with $h=\nu\circ h_x$ for all $x\in X$. Then $E(G,X)=E(G^{\mathrm{ad}},X^{\mathrm{ad}})\cdot E(T,h)$ (Deligne 1971*b*, 3.8).

(g) It follows from (e) and (f) that if (G,X) satisfies SV6, then $E(G,X)$ is either a totally real field or a CM-field.

Special points.

DEFINITION 12.5. A point $x\in X$ is said to be ***special*** if there exists a torus $T\subset G$ such that $h_x(\mathbb{C}^\times)\subset T(\mathbb{R})$. We then call (T,x), or (T,h_x), a ***special pair*** in (G,X). When the weight is rational and $Z(G)^\circ$ splits over a CM-field (i.e., SV4 and SV6 hold), the special points and special pairs are called ***CM points*** and ***CM pairs***.

REMARK 12.6. Let T be a maximal torus of G such that $T(\mathbb{R})$ fixes x, i.e., such that $\mathrm{ad}(t)\circ h_x=h_x$ for all $t\in T(\mathbb{R})$. Because $T_{\mathbb{R}}$ is its own centralizer in $G_{\mathbb{R}}$, this implies that $h_x(\mathbb{C}^\times)\subset T(\mathbb{R})$, and so x is special. Conversely, if (T,x) is special, then $T(\mathbb{R})$ fixes x.

EXAMPLE 12.7. Let $G=\mathrm{GL}_2$ and let $\mathcal{H}_1^{\pm}=\mathbb{C}\smallsetminus\mathbb{R}$. Then $G(\mathbb{R})$ acts on $\mathcal{H}_1^{\pm}$ by

$$\begin{pmatrix}a&b\\c&d\end{pmatrix}z=\frac{az+b}{cz+d}.$$

Suppose that $z\in\mathbb{C}\smallsetminus\mathbb{R}$ generates a quadratic imaginary extension E of $\mathbb{Q}$. Using the $\mathbb{Q}$-basis $\{1,z\}$ for E, we obtain an embedding $E\hookrightarrow M_2(\mathbb{Q})$, and hence a maximal subtorus $(\mathbb{G}_m)_{E/\mathbb{Q}}\subset G$. As $(\mathbb{G}_m)_{E/\mathbb{Q}}(\mathbb{R})$ fixes z, this shows that z is special. Conversely, if $z\in\mathcal{H}_1^{\pm}$ is special, then $\mathbb{Q}[z]$ is a field of degree 2 over $\mathbb{Q}$.

The homomorphism r_x. Let T be a torus over $\mathbb{Q}$ and let μ be a cocharacter of T defined over a finite extension E of $\mathbb{Q}$. For $Q\in T(E)$, the element $\sum_{\rho\colon E\to\mathbb{Q}^{\mathrm{al}}}\rho(Q)$ of $T(\mathbb{Q}^{\mathrm{al}})$ is stable under $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ and hence lies in $T(\mathbb{Q})$. Let $r(T,\mu)$ be the homomorphism $(\mathbb{G}_m)_{E/\mathbb{Q}}\to T$ such that

$$r(T,\mu)(P)=\sum_{\rho\colon E\to\mathbb{Q}^{\mathrm{al}}}\rho(\mu(P)),\quad\text{all }P\in E^\times.\tag{52}$$

Let $(T,x)\subset(G,X)$ be a special pair, and let $E(x)$ be the field of definition of μ_x. We define r_x to be the homomorphism

$$\mathbb{A}_{E(x)}^\times\xrightarrow{r(T,\mu)}T(\mathbb{A}_{\mathbb{Q}})\xrightarrow{\text{project}}T(\mathbb{A}_{\mathbb{Q},f}).\tag{53}$$

Let $a \in \mathbb{A}_{E(x)}^{\times}$, and write $a = (a_\infty, a_f) \in (E(x) \otimes_{\mathbb{Q}} \mathbb{R})^\times \times \mathbb{A}_{E(x),f}^{\times}$; then

$$r_x(a) = \sum_{\rho\colon E \to \mathbb{Q}^{\mathrm{al}}} \rho(\mu_x(a_f)).$$

Definition of a canonical model. For a special pair $(T, x) \subset (G, X)$, we have homomorphisms ((51),(53)),

$$\begin{aligned} \mathrm{art}_{E(x)}\colon \mathbb{A}_{E(x)}^{\times} &\twoheadrightarrow \mathrm{Gal}(E(x)^{\mathrm{ab}}/E(x)) \\ r_x\colon \mathbb{A}_{E(x)}^{\times} &\to T(\mathbb{A}_f). \end{aligned}$$

DEFINITION 12.8. Let (G, X) be a Shimura datum, and let K be a compact open subgroup of $G(\mathbb{A}_f)$. A model $M_K(G, X)$ of $\mathrm{Sh}_K(G, X)$ over $E(G, X)$ is ***canonical*** if, for every special pair $(T, x) \subset (G, X)$ and $a \in G(\mathbb{A}_f)$, $[x, a]_K$ has coordinates in $E(x)^{\mathrm{ab}}$ and

$$\sigma[x, a]_K = [x, r_x(s)a]_K, \tag{54}$$

for all

$$\left.\begin{array}{l} \sigma \in \mathrm{Gal}(E(x)^{\mathrm{ab}}/E(x)) \\ s \in \mathbb{A}_{E(x)}^{\times} \end{array}\right\} \text{ with } \mathrm{art}_{E(x)}(s) = \sigma.$$

In other words, $M_K(G, X)$ is canonical if every automorphism σ of $\mathbb{C}$ fixing $E(x)$ acts on $[x, a]_K$ according to the rule (54) where s is any idèle such that $\mathrm{art}_{E(x)}(s) = \sigma|E(x)^{\mathrm{ab}}$.

REMARK 12.9. Let (T_1, x) and (T_2, x) be special pairs in (G, X) (with the same x). Then $(T_1 \cap T_2, x)$ is also a special pair, and if the condition in (54) holds for one of $(T_1 \cap T_2, x)$, (T_1, x), or (T_2, x), then it holds for all three. Therefore, in stating the definition, we could have considered only special pairs (T, x) with, for example, T minimal among the tori such that $T_{\mathbb{R}}$ contains $h_x(\mathbb{S})$.

DEFINITION 12.10. Let (G, X) be a Shimura datum.

(a) A ***model*** of $\mathrm{Sh}(G, X)$ over a subfield k of $\mathbb{C}$ is an inverse system $M(G, X) = (M_K(G, X))_K$ of varieties over k endowed with a right action of $G(\mathbb{A}_f)$ such that $M(G, X)_{\mathbb{C}} = \mathrm{Sh}(G, X)$ (with its $G(\mathbb{A}_f)$ action).

(b) A model $M(G, X)$ of $\mathrm{Sh}(G, X)$ over $E(G, X)$ is ***canonical*** if each $M_K(G, X)$ is canonical.

Examples: Shimura varieties defined by tori. For a field k of characteristic zero, the functor $V \mapsto V(k^{\mathrm{al}})$ is an equivalence from the category of zero-dimensional varieties over k to the category of finite sets endowed with a continuous action of $\mathrm{Gal}(k^{\mathrm{al}}/k)$. Continuous here just means that the action factors through $\mathrm{Gal}(L/k)$ for some finite galois extension L of k contained in k^{al}. In particular, to give a zero-dimensional variety over an algebraically closed field of characteristic zero is just to give a finite set. Thus, a zero-dimensional variety over $\mathbb{C}$ can be regarded as a zero-dimensional variety over $\mathbb{Q}^{\mathrm{al}}$, and to give a model of V over a number field E amounts to giving a continuous action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ on $V(\mathbb{C})$.

Tori. Let T be a torus over $\mathbb{Q}$, and let h be a homomorphism $\mathbb{S} \to T_{\mathbb{R}}$. Then (T, h) is a Shimura datum, and $E =_{\mathrm{df}} E(T, h)$ is the field of definition of μ_h. In this case

$$\mathrm{Sh}_K(T, h) = T(\mathbb{Q}) \backslash \{h\} \times T(\mathbb{A}_f)/K$$

is a finite set (see 5.22), and (54) defines a continuous action of $\mathrm{Gal}(E^{\mathrm{ab}}/E)$ on $\mathrm{Sh}_K(T,h)$. This action defines a model of $\mathrm{Sh}_K(T,h)$ over E, which, by definition, is canonical.

CM-tori. Let (E,Φ) be a CM-type, and let (T,h_Φ) be the Shimura pair defined in (12.4b). Then $E(T,h_\Phi)=E^*$, and $r(T,\mu_\Phi)\colon (\mathbb{G}_m)_{E^*/\mathbb{Q}}\to(\mathbb{G}_m)_{E/\mathbb{Q}}$ is the reflex norm N_{Φ^*}.

Let K be a compact open subgroup of $T(\mathbb{A}_f)$. The Shimura variety $\mathrm{Sh}_K(T,h_\Phi)$ classifies isomorphism classes of triples $(A,i,\eta K)$ in which (A,i) is an abelian variety over $\mathbb{C}$ of CM-type (E,Φ) and η is an $E\otimes\mathbb{A}_f$-linear isomorphism $V(\mathbb{A}_f)\to V_f(A)$. An isomorphism $(A,i,\eta K)\to(A',i',\eta' K)$ is an E-linear isomorphism $A\to A'$ in $\mathsf{AV}^0(\mathbb{C})$ sending ηK to $\eta' K$. To see this, let V be a one-dimensional E-vector space. The action of E on V realizes T as a subtorus of $\mathrm{GL}(V)$. If (A,i) is of CM-type (E,Φ), then there exists an E-homomorphism $a\colon H_1(A,\mathbb{Q})\to V$ carrying h_A to h_Φ (see 10.2). Now the isomorphism

$$V(\mathbb{A}_f)\xrightarrow{\eta}V_f(A)\xrightarrow{a}V(\mathbb{A}_f)$$

is $E\otimes\mathbb{A}_f$-linear, and hence is multiplication by an element g of $(E\otimes\mathbb{A}_f)^\times=T^E(\mathbb{A}_f)$. The map $(A,i,\eta)\mapsto[g]$ gives the bijection.

In (10.3) and its proof, we showed that the functor $(A,i)\mapsto(A_{\mathbb{C}},i_{\mathbb{C}})$ defines an equivalence from the category of abelian varieties over $\mathbb{Q}^{\mathrm{al}}$ of CM-type (E,Φ) to the similar category over $\mathbb{C}$ (the abelian varieties are to be regarded as objects of AV^0). Therefore, $\mathrm{Sh}_K(T^E,h_\Phi)$ classifies isomorphism classes of triples $(A,i,\eta K)$ where (A,i) is now an abelian variety over $\mathbb{Q}^{\mathrm{al}}$ of CM-type (E,Φ).

The group $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*)$ acts on the set $\mathcal{M}_K$ of such triples: let $(A,i,\eta)\in\mathcal{M}_K$; for $\sigma\in\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*)$, define $\sigma(A,i,\eta K)$ to be the triple $(\sigma A,{}^\sigma i,{}^\sigma\eta K)$ where ${}^\sigma\eta$ is the composite

$$V(\mathbb{A}_f)\xrightarrow{\eta}V_f(A)\xrightarrow{\sigma}V_f(\sigma A);\tag{55}$$

because σ fixes E^*, $(\sigma A,\sigma i)$ is again of CM-type (E,Φ).

The group $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*)$ acts on $\mathrm{Sh}_K(T^E,h_\Phi)$ by the rule (54):

$$\sigma[g]=[r_{h_\Phi}(s)g]_K,\quad \mathrm{art}_{E^*}(s)=\sigma|E^*.$$

PROPOSITION 12.11. *The map* $(A,i,\eta)\mapsto[a\circ\eta]_K\colon\mathcal{M}_K\to\mathrm{Sh}_K(T^E,h_\Phi)$ *commutes with the actions of* $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*)$.

PROOF. Let $(A,i,\eta)\in\mathcal{M}_K$ map to $[a\circ\eta]_K$ for an appropriate isomorphism $a\colon H_1(A,\mathbb{Q})\to V$, and let $\sigma\in\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*)$. According to the main theorem of complex multiplication (11.2), there exists an isomorphism $\alpha\colon A\to\sigma A$ such that $\alpha(N_{\Phi^*}(s)\cdot x)=\sigma x$ for $x\in V_f(A)$, where $s\in\mathbb{A}_{E^*}$ is such that $\mathrm{art}_{E^*}(s)=\sigma|E^*$. Then $\sigma(A,i,\eta)\mapsto[a\circ H_1(\alpha)^{=1}\circ\sigma\circ\eta]_K$. But

$$V_f(\alpha)^{=1}\circ\sigma=N_{\Phi^*}(s)=r_{h_\Phi}(s),$$

and so

$$[a\circ H_1(\alpha)^{-1}\circ\sigma\circ\eta]_K=[r_{h_\Phi}(s)\cdot(a\circ\eta)]_K$$

as required. □

NOTES. Our definitions coincide with those of Deligne 1979, except that we have corrected a sign error there (it is necessary to delete "*inverse*" in ibid. 2.2.3, p269, line 10, and in 2.6.3, p284, line 21).

13. Uniqueness of canonical models

In this section, I sketch a proof that a Shimura variety has at most one canonical model (up to a unique isomorphism).

Extension of the base field.

PROPOSITION 13.1. *Let k be a subfield of an algebraically closed field Ω of characteristic zero. If V and W are varieties over k, then a regular map $V_\Omega \to W_\Omega$ commuting with the actions of $\mathrm{Aut}(\Omega/k)$ on $V(\Omega)$ and $W(\Omega)$ arises from a unique regular map $V \to W$. In other words, the functor*

$$V \mapsto V_\Omega + \textit{action of } \mathrm{Aut}(\Omega/k) \textit{ on } V(\Omega)$$

is fully faithful.

PROOF. See AG 16.9. [The first step is to show that the $\Omega^{\mathrm{Aut}(\Omega/k)} = k$, which requires Zorn's lemma in general.] □

COROLLARY 13.2. *A variety V over k is uniquely determined (up to a unique isomorphism) by V_Ω and the action of $\mathrm{Aut}(\Omega/k)$ on $V(\Omega)$.*

Uniqueness of canonical models. Let (G, X) be a Shimura datum.

LEMMA 13.3. *There exists a special point in X.*

PROOF (SKETCH). Let $x \in X$, and let T be a maximal torus in $G_\mathbb{R}$ containing $h_x(\mathbb{C})$. Then T is the centralizer of any regular element λ of $\mathrm{Lie}(T)$. If $\lambda_0 \in \mathrm{Lie}(G)$ is chosen sufficiently close to λ, then the centralizer T_0 of λ_0 in G will be a maximal torus in G (Borel 1991, 18.1, 18.2), and T_0 will become conjugate[17] to T over $\mathbb{R}$:

$$T_{0\mathbb{R}} = gTg^{-1}, \text{ some } g \in G(\mathbb{R}).$$

Now $h_{gx}(\mathbb{S}) =_{\mathrm{df}} ghg^{-1}(\mathbb{S}) \subset T_{0\mathbb{R}}$, and so gx is special. □

LEMMA 13.4 (Key Lemma). *For any finite extension L of $E(G, X)$ in $\mathbb{C}$, there exists a special point x_0 such that $E(x_0)$ is linearly disjoint from L.*

PROOF. See Deligne 1971*b*, 5.1. [The basic idea is the same as that of the proof of 13.3 above, but requires the Hilbert irreducibility theorem.] □

If $G = \mathrm{GL}_2$, the lemma just says that, for any finite extension L of $\mathbb{Q}$ in $\mathbb{C}$, there exists a quadratic imaginary extension E over $\mathbb{Q}$ linearly disjoint from L. This is obvious — for example, take $E = \mathbb{Q}[\sqrt{-p}]$ for any prime p unramified in L.

LEMMA 13.5. *For any $x \in X$, $\{[x, a]_K \mid a \in G(\mathbb{A}_f)\}$ is dense in $\mathrm{Sh}_K(G, X)$ (in the zariski topology).*

PROOF. Write

$$\mathrm{Sh}_K(G, X)(\mathbb{C}) = G(\mathbb{Q})\backslash X \times (G(\mathbb{A}_f)/K)$$

and note that the real approximation theorem (5.4) implies that $G(\mathbb{Q})x$ is dense in X for the complex topology, and, a fortiori, the zariski topology. □

[17] Any element sufficiently close to a regular element will also be regular, which implies that T_0 is a maximal torus. Not all maximal tori in $G_{/\mathbb{R}}$ are conjugate — rather, they fall into several connected components, from which the second statement can be deduced.

Let $g \in G(\mathbb{A}_f)$, and let K and K' be compact open subgroups such that $K' \supset g^{-1}Kg$. Then the map $\mathcal{T}(g)$

$$[x,a]_K \mapsto [x,ag]_{K'} \colon \operatorname{Sh}_K(\mathbb{C}) \to \operatorname{Sh}_{K'}(\mathbb{C})$$

is well-defined.

THEOREM 13.6. *If* $\operatorname{Sh}_K(G,X)$ *and* $\operatorname{Sh}_{K'}(G,X)$ *have canonical models over* $E(G,X)$*, then* $\mathcal{T}(g)$ *is defined over* $E(G,X)$.

PROOF. After (13.1), it suffices to show that $\sigma(\mathcal{T}(g)) = \mathcal{T}(g)$ for all automorphisms σ of $\mathbb{C}$ fixing $E(G,X)$. Let $x_0 \in X$ be special. Then $E(x_0) \supset E(G,X)$ (see 12.3b), and we first show that $\sigma(\mathcal{T}(g)) = \mathcal{T}(g)$ for those σ's fixing $E(x_0)$. Choose an $s \in \mathbb{A}_{E_0}^\times$ such that $\operatorname{art}(s) = \sigma|E(x_0)^{\mathrm{ab}}$. For $a \in G(\mathbb{A}_f)$,

$$\begin{array}{ccc} [x_0,a]_K & \xmapsto{\mathcal{T}(g)} & [x_0,ag]_{K'} \\ \Big\downarrow{\scriptstyle\sigma} & & \Big\downarrow{\scriptstyle\sigma} \\ [x_0, r_{x_0}(s)a]_K & \xmapsto{\mathcal{T}(g)} & [x_0, r_{x_0}(s)ag]_{K'} \end{array}$$

commutes. Thus, $\mathcal{T}(g)$ and $\sigma(\mathcal{T}(g))$ agree on $\{[x_0,a] \mid a \in G(\mathbb{A}_f)\}$, and hence on all of Sh_K by Lemma 13.5. We have shown that $\sigma(\mathcal{T}(g)) = \mathcal{T}(g)$ for all σ fixing the reflex field of any special point, but Lemma 13.4 shows that these σ's generate $\operatorname{Aut}(\mathbb{C}/E(G,X))$. □

THEOREM 13.7. *(a) A canonical model of* $\operatorname{Sh}_K(G,X)$ *(if it exists) is unique up to a unique isomorphism.*

(b) If, for all compact open subgroups K of $G(\mathbb{A}_f)$, $\operatorname{Sh}_K(G,X)$ *has a canonical model, then so also does* $\operatorname{Sh}(G,X)$*, and it is unique up to a unique isomorphism.*

PROOF. (a) Take $K = K'$ and $g = 1$ in (13.6).

(b) Obvious from (13.6). □

In more detail, let $(M_K(G,X),\varphi)$ and $(M'_K(G,X),\varphi')$ be canonical models of $\operatorname{Sh}_K(G,X)$ over $E(G,X)$. Then the composite

$$M_K(G,X)_{\mathbb{C}} \xrightarrow{\varphi} \operatorname{Sh}_K(G,X) \xrightarrow{\varphi'^{-1}} M'_K(G,X)_{\mathbb{C}}$$

is fixed by all automorphisms of $\mathbb{C}$ fixing $E(G,X)$, and is therefore defined over $E(G,X)$.

REMARK 13.8. In fact, one can prove more. Let $a\colon (G,X) \to (G',X')$ be a morphism of Shimura data, and suppose $\operatorname{Sh}(G,X)$ and $\operatorname{Sh}(G',X')$ have canonical models $M(G,X)$ and $M(G',X')$. Then the morphism $\operatorname{Sh}(a)\colon \operatorname{Sh}(G,X) \to \operatorname{Sh}(G',X')$ is defined over $E(G,X) \cdot E(G',X')$.

The galois action on the connected components. A canonical model for $\operatorname{Sh}_K(G,X)$ will define an action of $\operatorname{Aut}(\mathbb{C}/E(G,X))$ on the set $\pi_0(\operatorname{Sh}_K(G,X))$. In the case that G^{der} is simply connected, we saw in §5 that

$$\pi_0(\operatorname{Sh}_K(G,X)) \cong T(\mathbb{Q})\backslash Y \times T(\mathbb{A}_f)/\nu(K)$$

where $\nu\colon G \to T$ is the quotient of G by G^{der} and Y is the quotient of $T(\mathbb{R})$ by the image $T(\mathbb{R})^\dagger$ of $Z(\mathbb{R})$ in $T(\mathbb{R})$. Let $h = \nu \circ h_x$ for any $x \in X$. Then μ_h is certainly defined over $E(G,X)$. Therefore, it defines a homomorphism

$$r = r(T,\mu_h)\colon \mathbb{A}_{E(G,X)}^\times \to T(\mathbb{A}_{\mathbb{Q}}).$$

The action of $\sigma \in \mathrm{Aut}(\mathbb{C}/E(G,X))$ on $\pi_0(\mathrm{Sh}_K(G,X))$ can be described as follows: let $s \in \mathbb{A}^\times_{E(G,X)}$ be such that $\mathrm{art}_{E(G,X)}(s) = \sigma|E(G,X)^{\mathrm{ab}}$, and let $r(s) = (r(s)_\infty, r(s)_f) \in T(\mathbb{R}) \times T(\mathbb{A}_f)$; then

$$\sigma[y,a]_K = [r(s)_\infty y, r(s)_f \cdot a]_K, \text{ for all } y \in Y, \quad a \in T(\mathbb{A}_f). \tag{56}$$

When we use (56) to define the notion a canonical model of a zero-dimensional Shimura variety, we can say that π_0 of the canonical model of $\mathrm{Sh}_K(G,X)$ is the canonical model of $\mathrm{Sh}(T,Y)$.

If σ fixes a special x_0 mapping to y, then (56) follows from (54), and a slight improvement of (13.4) shows that such σ's generate $\mathrm{Aut}(\mathbb{C}/E(G,X))$.

NOTES. The proof of uniqueness follows Deligne 1971*b*, §3, except that I am more unscrupulous in my use of the Zorn's lemma.

14. Existence of canonical models

Canonical models are known to exist for all Shimura varieties. In this section, I explain some of the ideas that go into the proof.

Descent of the base field. Let k be a subfield of an algebraically closed field Ω of characteristic zero, and let $\mathcal{A} = \mathrm{Aut}(\Omega/k)$. In (13.1) we observed that the functor

$$\{\text{varieties over } k\} \to \{\text{varieties } V \text{ over } \Omega + \text{action of } \mathcal{A} \text{ on } V(\Omega)\},$$

is fully faithful. In this subsection, we find conditions on a pair $(V, \cdot)$ that ensure that it is in the essential image of the functor, i.e., that it arises from a variety over k. We begin by listing two necessary conditions.

The regularity condition. Obviously, the action $\cdot$ should recognize that $V(\Omega)$ is not just a set, but rather the set of points of an algebraic variety. Recall that, for $\sigma \in \mathcal{A}$, σV is obtained from V by applying σ to the coefficients of the polynomials defining V, and $\sigma P \in (\sigma V)(\Omega)$ is obtained from $P \in V(\Omega)$ by applying σ to the coordinates of P.

DEFINITION 14.1. An action $\cdot$ of $\mathcal{A}$ on $V(\Omega)$ is ***regular*** if the map

$$\sigma P \mapsto \sigma \cdot P \colon (\sigma V)(\Omega) \to V(\Omega)$$

is a regular isomorphism for all σ.

A priori, this is only a map of sets. The condition requires that it be induced by a regular map $f_\sigma \colon \sigma V \to V$. If $(V, \cdot)$ arises from a variety over k, then $\sigma V = V$ and $\sigma P = \sigma \cdot P$, and so the condition is clearly necessary.

REMARK 14.2. (a) When regular, the maps f_σ are automatically isomorphisms provided V is nonsingular.

(b) The maps f_σ satisfy the cocycle condition $f_\sigma \circ \sigma f_\tau = f_{\sigma\tau}$. Conversely, every family $(f_\sigma)_{\sigma \in \mathcal{A}}$ of regular isomorphisms satisfying the cocycle condition arises from an action of $\mathcal{A}$ satisfying the regularity condition. Such families $(f_\sigma)_{\sigma \in \mathcal{A}}$ are called ***descent data***, and normally one expresses descent theory in terms of them rather than actions of $\mathcal{A}$.

The continuity condition.

DEFINITION 14.3. An action $\cdot$ of $\mathcal{A}$ on $V(\Omega)$ is ***continuous*** if there exists a subfield L of Ω finitely generated over k and a model V_0 of V over L such that the action of $\mathrm{Aut}(\Omega/L)$ on $V(\Omega)$ defined by V_0 is $\cdot$.

More precisely, the condition requires that there exist a model (V_0, φ) of V over L such that $\varphi(\sigma P) = \sigma \cdot \varphi(P)$ for all $P \in V_0(\Omega)$ and $\sigma \in \mathrm{Aut}(\mathbb{C}/L)$. Clearly this condition is necessary.

PROPOSITION 14.4. *A regular action $\cdot$ of $\mathcal{A}$ on $V(\Omega)$ is continuous if there exist points $P_1, \ldots, P_n \in V(\Omega)$ such that*

(a) *the only automorphism of V fixing every P_i is the identity map;*
(b) *there exists a subfield L of Ω finitely generated over k such that $\sigma \cdot P_i = P_i$ for all σ fixing L.*

PROOF. Let (V_0, φ) be a model of V over a subfield L of Ω finitely generated over k. After possibly enlarging L, we may assume that $\varphi^{-1}(P_i) \in V_0(L)$ and that $\sigma \cdot P_i = P_i$ for all σ fixing L (because of (b)). For such a σ, f_σ and $\varphi \circ (\sigma\varphi)^{-1}$ are regular maps $\sigma V \to V$ sending σP_i to P_i for each i, and so they are equal (because of (a)). Hence

$$\varphi(\sigma P) = f_\sigma((\sigma\varphi)(\sigma P)) = f_\sigma(\sigma(\varphi(P))) = \sigma \cdot \varphi(P)$$

for all $P \in V_0(\Omega)$, and so the action of $\mathrm{Aut}(\mathbb{C}/L)$ on $V(\Omega)$ defined by (V_0, φ) is $\cdot$. □

A sufficient condition for descent.

THEOREM 14.5. *If V is quasiprojective and $\cdot$ is regular and continuous, then $(V, \cdot)$ arises from a variety over k.*

PROOF. This is a restatement of the results of Weil 1956 (see Milne 1999, 1.1). □

COROLLARY 14.6. *The pair $(V, \cdot)$ arises from a variety over k if*

(a) *V is quasiprojective,*
(b) *$\cdot$ is regular, and*
(c) *there exists points $P_1, \ldots, P_n$ in $V(\Omega)$ satisfying the conditions (a) and (b) of (14.4).*

PROOF. Immediate from (14.5) and (14.6). □

For an elementary proof of the corollary, not using the results of Weil 1956, see AG 16.33.

Review of local systems and families of abelian varieties. Let S be a topological manifold. A ***local system of $\mathbb{Z}$-modules on*** S is a sheaf F on S that is locally isomorphic to the constant sheaf $\mathbb{Z}^n$ $(n \in \mathbb{N})$.

Let F be a local system of $\mathbb{Z}$-modules on S, and let $o \in S$. There is an action of $\pi_1(S, o)$ on F_o that can be described as follows: let $\gamma\colon [0,1] \to S$ be a loop at o; because $[0,1]$ is simply connected, there is an isomorphism from γ^*F to the constant sheaf defined by a group M say; when we choose such an isomorphism, we obtain isomorphisms $(\gamma^*F)_i \to M$ for all $i \in [0,1]$; now $(\gamma^*F)_i = F_{\gamma(i)}$ and $\gamma(0) = o = \gamma(1)$, and so we get two isomorphisms $F_o \to M$; these isomorphisms differ by an automorphism of F_o, which depends only the homotopy class of γ.

PROPOSITION 14.7. *If S is connected, then $F \mapsto (F_o, \rho_o)$ defines an equivalence from the category of local systems of $\mathbb{Z}$-modules on S to the category of finitely generated $\mathbb{Z}$-modules endowed with an action of $\pi_1(S, o)$.*

PROOF. This is well known; cf. Deligne 1970, I 1. □

Let F be a local system of $\mathbb{Z}$-modules on S. Let $\pi\colon \tilde{S} \to S$ be the universal covering space of S, and choose a point $o \in \tilde{S}$. We can identifiy π^*F with the constant sheaf defined by $F_{\pi(o)}$. Suppose that we have a hodge structure h_s on $F_s \otimes \mathbb{R}$ for every $s \in S$. We say that F, together with the hodge structures, is a ***variation of integral hodge structures on*** S if $s \mapsto h_{\pi(s)}$ (hodge structure on $F_{\pi(o)} \otimes \mathbb{R}$) is a variation of hodge structures on $\tilde{S}$. A ***polarization*** of a variation of hodge structures $(F, (h_s))$ is a pairing $\psi\colon F \times F \to \mathbb{Z}$ such that ψ_s is a polarization of (F_s, h_s) for every s.

Let V be a nonsingular algebraic variety over $\mathbb{C}$. A ***family of abelian varieties over*** V is a regular map $f\colon A \to V$ of nonsingular varieties plus a regular multiplication $A \times_V A \to A$ over V such that the fibres of f are abelian varieties of constant dimension (in a different language, A is an abelian scheme over V).

THEOREM 14.8. *Let V be a nonsingular variety over $\mathbb{C}$. There is an equivalence $(A, f) \mapsto (R^1 f_*\mathbb{Z})^\vee$ from the category of families of abelian varieties over V to the category of polarizable integral variations of hodge structures of type $(-1,0)$, $(0,-1)$ on S.*

This is a generalization of Riemann's theorem (6.8) — see Deligne 1971*a*, 4.4.3.

The Siegel modular variety. Let (V, ψ) be a symplectic space over $\mathbb{Q}$, and let $(G, X) = (\mathrm{GSp}(\psi), X(\psi))$ be the associated Shimura datum (§6). We also denote $\mathrm{Sp}(\psi)$ by S. We abbreviate $\mathrm{Sh}_K(G, X)$ to Sh_K.

The reflex field. Consider the set of pairs (L, L') of complementary lagrangians in $V(\mathbb{C})$:

$$V(\mathbb{C}) = L \oplus L', \quad L, L' \text{ totally isotropic.} \tag{57}$$

Every symplectic basis for $V(\mathbb{C})$ defines such a pair, and every such pair arises from a symplectic basis. Therefore, $G(\mathbb{C})$ (even $S(\mathbb{C})$) acts transitively on the set of pairs (L, L') of complementary lagrangians. For such a pair, let $\mu_{(L,L')}$ be the homomorphism $\mathbb{G}_m \to \mathrm{GL}(V)$ such that $\mu(z)$ acts as z on L and as 1 on L'. Then, $\mu_{(L,L')}$ takes values in $G_{\mathbb{C}}$, and as (L, L') runs through the set of pairs of complementary lagrangians in $V(\mathbb{C})$, $\mu_{(L,L')}$ runs through $c(X)$ (notation as on p343). Since V itself has symplectic bases, there exist pairs of complementary lagrangians in V. For such a pair, $\mu_{(L,L')}$ is defined over $\mathbb{Q}$, and so $c(X)$ has a representative defined over $\mathbb{Q}$. This shows that the reflex field $E(G, X) = \mathbb{Q}$.

The special points. Let K be a compact open subgroup of $G(\mathbb{A}_f)$, and, as in §6, let $\mathcal{M}_K$ be the set of triples $(A, s, \eta K)$ in which A is an abelian variety over $\mathbb{C}$, s is an alternating form on $H_1(A, \mathbb{Q})$ such that $\pm s$ is a polarization, and η is an isomorphism $V(\mathbb{A}_f) \to V_f(A)$ sending ψ to a multiple of s. Recall (6.11) that there is a natural map $\mathcal{M}_K \to \mathrm{Sh}_K(\mathbb{C})$ whose fibres are the isomorphism classes.

In this subsubsection we answer the question: which triples $(A, s, \eta K)$ correspond to points $[x, a]$ with x special?

DEFINITION 14.9. A ***CM-algebra*** is a finite product of CM-fields. An abelian variety A over $\mathbb{C}$ is ***CM*** if there exists a CM-algebra E and a homomorphism $E \to \mathrm{End}^0(A)$ such that $H_1(A, \mathbb{Q})$ is a free E-module of rank 1.

Let $E \to \mathrm{End}^0(A)$ be as in the definition, and let E be a product of CM-fields $E_1, \ldots, E_m$. Then A is isogenous to a product of abelian varieties $A_1 \times \cdots \times A_m$ with A_i of CM-type (E_i, Φ_i) for some Φ_i.

Recall that, for an abelian variety A over $\mathbb{C}$, there is a homomorphism $h_A \colon \mathbb{C}^\times \to \mathrm{GL}(H_1(A, \mathbb{R}))$ describing the natural complex structure on $H_1(A, \mathbb{R})$ (see §6).

PROPOSITION 14.10. *An abelian variety A over $\mathbb{C}$ is CM if and only if there exists a torus $T \subset \mathrm{GL}(H_1(A, \mathbb{Q}))$ such that $h_A(\mathbb{C}^\times) \subset T(\mathbb{R})$.*

PROOF. See Mumford 1969, §2, or Deligne 1982, §3. □

COROLLARY 14.11. *If $(A, s, \eta K) \mapsto [x, a]_K$ under $\mathcal{M}_K \to \mathrm{Sh}_K(G, X)$, then A is of CM-type if and only if x is special.*

PROOF. Recall that if $(A, s, \eta K) \mapsto [x, a]_K$, then there exists an isomorphism $H_1(A, \mathbb{Q}) \to V$ sending h_A to h_x. Thus, the statement follows from the proposition. □

A criterion to be canonical. We now define an action of $\mathrm{Aut}(\mathbb{C})$ on $\mathcal{M}_K$. Let $(A, s, \eta K) \in \mathcal{M}_K$. Then $s \in H^2(A, \mathbb{Q})$ is a hodge tensor, and therefore equals $r[D]$ for some $r \in \mathbb{Q}^\times$ and divisor D on A (see 7.5). We let ${}^\sigma s = r[\sigma D]$. The condition that $\pm s$ be positive definite is equivalent to an algebro-geometric condition on D (Mumford 1970, pp29–30) which is preserved by σ. Therefore, $\pm {}^\sigma s$ is a polarization for $H_1(A, \mathbb{Q})$. We define $\sigma(A, s, \eta K)$ to be $(\sigma A, {}^\sigma s, {}^\sigma \eta K)$ with ${}^\sigma \eta$ as in (55).

PROPOSITION 14.12. *Suppose that Sh_K has a model M_K over $\mathbb{Q}$ for which the map*

$$\mathcal{M}_K \to M_K(\mathbb{C})$$

commutes with the actions of $\mathrm{Aut}(\mathbb{C})$. Then M_K is canonical.

PROOF. For a special point $[x, a]_K$ corresponding to an abelian variety A with complex multiplication by a field E, the condition (54) is an immediate consequence of the main theorem of complex multiplication (cf. 12.11). For more general special points, it also follows from the main theorem of complex multiplication, but not quite so immediately. □

Outline of the proof of the existence of a canonical model. Since the action of $\mathrm{Aut}(\mathbb{C})$ on $\mathcal{M}_K$ preserves the isomorphism classes, from the map $\mathcal{M}_K \to \mathrm{Sh}_K(\mathbb{C})$, we get an action of $\mathrm{Aut}(\mathbb{C})$ on $\mathrm{Sh}_K(\mathbb{C})$. If this action satisfies the conditions of hypotheses of Corollary 14.6, then $\mathrm{Sh}_K(G, X)$ has a model over $\mathbb{Q}$, which Proposition 14.12 will show to be canonical.

Condition (a) of (14.6). We know that $\mathrm{Sh}_K(G, X)$ is quasi-projective from (3.12).

Condition (b) of (14.6). We have to show that the map

$$\sigma P \mapsto \sigma \cdot P \colon \sigma\,\mathrm{Sh}_K(\mathbb{C}) \xrightarrow{f_\sigma} \mathrm{Sh}_K(\mathbb{C})$$

is regular. It suffices to do this for K small, because if $K' \supset K$, then $\mathrm{Sh}_{K'}(G, X)$ is a quotient of $\mathrm{Sh}_K(G, X)$.

Recall (5.17) that $\pi_0(\mathrm{Sh}_K) \cong \mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$. Let $\varepsilon \in \mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$, and let $\mathrm{Sh}_K^\varepsilon$ be the corresponding connected component of Sh_K. Then $\mathrm{Sh}_K^\varepsilon = \Gamma_\varepsilon\backslash X^+$ where $\Gamma_\varepsilon = G(\mathbb{Q}) \cap K_\varepsilon$ for some conjugate K_ε of K (see 5.17, 5.23)

Let $(A, s, \eta K) \in \mathcal{M}_K$ and choose an isomorphism $a\colon H_1(A,\mathbb{Q}) \to V$ sending s to a multiple of ψ. Then the image of $(A, s, \eta K)$ in $\mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$ is represented by $\nu(a\circ\eta)$ where $a\circ\eta\colon V(\mathbb{A}_f) \to V(\mathbb{A}_f)$ is to be regarded as an element of $G(\mathbb{A}_f)$. Write $\mathcal{M}_K^\varepsilon$ for the set of triples with $\nu(a\circ\eta) \in \varepsilon$. Define $\mathcal{H}_K^\varepsilon$ similarly.

The map $\mathcal{M}_K \to \mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$ is equivariant for the action of $\mathrm{Aut}(\mathbb{C})$ when we let $\mathrm{Aut}(\mathbb{C})$ act on $\mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$ through the cyclotomic character, i.e.,

$$\sigma[\alpha] = [\chi(\sigma)\alpha] \text{ where } \chi(\sigma) \in \hat{\mathbb{Z}}^\times,\ \zeta^{\chi(\sigma)} = \sigma\zeta,\ \zeta \text{ a root of 1.}$$

Write $X^+(\Gamma_\varepsilon)$ for $\Gamma_\varepsilon\backslash X^+$ regarded as an algebraic variety, and let $\sigma(X^+(\Gamma_\varepsilon))$ be the algebraic variety obtained from $X^+(\Gamma_\varepsilon)$ by change of base field $\sigma\colon \mathbb{C} \to \mathbb{C}$. Consider the diagram:

$$\begin{array}{ccc}
X^+ & \xleftarrow{\alpha} & U \\
\downarrow & & \downarrow \\
X^+(\Gamma_{\sigma\varepsilon}) & \xleftarrow{f_\sigma} & \sigma(X^+(\Gamma_\varepsilon)) \\
\mathcal{M}_K^{\sigma\varepsilon} & \xleftarrow{\sigma} & \mathcal{M}_K^\varepsilon
\end{array}$$

The map σ sends $(A,\ldots)$ to $\sigma(A,\ldots)$, and the map f_σ is the map of sets $\sigma P \mapsto \sigma\cdot P$. The two maps are compatible. The map $U \to \sigma(X^+(\Gamma_\varepsilon))$ is the universal covering space of the complex manifold $(\sigma(X^+(\Gamma_\varepsilon)))^{\mathrm{an}}$.

Fix a lattice Λ in V that is stable under the action of Γ_ε. From the action of Γ_ε on Λ, we get a local system of $\mathbb{Z}$-modules M on $X^+(\Gamma_\varepsilon)$ (see 14.7), which, in fact, is a polarized integral variation of hodge structures F. According to Theorem 14.8, this variation of hodge structures arises from a polarized family of abelian varieties $f\colon \mathcal{A} \to X^+(\Gamma_\varepsilon)$. As f is a regular map of algebraic varieties, we can apply σ to it, and obtain a polarized family of abelian varieties $\sigma f\colon \sigma\mathcal{A} \to \sigma(X^+(\Gamma_\varepsilon))$. Then $(R^1(\sigma f)_*\mathbb{Z})^\vee$ is a polarized integral hodge structure on $\sigma(X^+(\Gamma_\varepsilon))$. On pulling this back to U and tensoring with $\mathbb{Q}$, we obtain a variation of polarized rational hodge structures over the space U, whose underlying local system can identified with the constant sheaf defined by V. When this identification is done correctly, each $u \in U$ defines a complex structure on V that is positive for ψ, i.e., a point x of X^+, and the map $u \mapsto x$ makes the diagram commute. Now (2.15) shows that $u \mapsto x$ is holomorphic. It follows that f_σ is holomorphic, and Borel's theorem (3.14) shows that it is regular.

Condition (c) of (14.6). For any $x \in X$, the set $\{[x,a]_K \mid a \in G(\mathbb{A}_f)\}$ has the property that only the identity automorphism of $\mathrm{Sh}_K(G,X)$ fixes its elements (see 13.5). But, there are only finitely many automorphisms of $\mathrm{Sh}_K(G,X)$ (see 3.21), and so a finite sequence of points $[x,a_1],\ldots,[x,a_n]$ will have this property. When we choose x to be special, the main theorem of complex multiplication (11.2) tells us that $\sigma\cdot[x,a_i] = [x,a_i]$ for all σ fixing some fixed finite extension of $E(x)$, and so condition (c) holds for these points.

Simple PEL Shimura varieties of type A or C. The proof is similar to the Siegel case. Here $\mathrm{Sh}_K(G,X)$ classifies quadruples $(A,i,s,\eta K)$ satisfying certain conditions. One checks that if σ fixes the reflex field $E(G,X)$, then $\sigma(A,i,s,\eta K)$

lies in the family again (see 12.7). Again the special points correspond to CM abelian varieties, and the Shimura-Taniyama theorem shows that, if $\mathrm{Sh}_K(G,X)$ has a model M_K over $E(G,X)$ for which the action of $\mathrm{Aut}(\mathbb{C}/E(G,X))$ on $M_K(\mathbb{C}) = \mathrm{Sh}_K(G,X)(\mathbb{C})$ agrees with its action on the quadruples, then it is canonical.

Shimura varieties of hodge type. In this case, $\mathrm{Sh}_K(G,X)$ classifies isomorphism classes of triples $(A,(s_i)_{0\le i\le n},\eta K)$ where the s_i are hodge tensors. A proof similar to that in the Siegel case will apply once we have defined ${}^{\sigma}s$ for s a hodge tensor on an abelian variety.

If the Hodge conjecture is true, then s is the cohomology class of some algebraic cycle Z on A (i.e., formal $\mathbb{Q}$-linear combination of integral subvarieties of A). Then we could define ${}^{\sigma}s$ to be the cohomology class of σZ on σA. Unfortunately, a proof of the Hodge conjecture seems remote, even for abelian varieties. Deligne succeeded in defining ${}^{\sigma}s$ without the Hodge conjecture. It is important to note that there is no natural map between $H^n(A,\mathbb{Q})$ and $H^n(\sigma A,\mathbb{Q})$ (unless σ is continuous, and hence is the identity or complex conjugation). However, there is a natural isomorphism $\sigma\colon H^n(A,\mathbb{A}_f)\to H^n(\sigma A,\mathbb{A}_f)$ coming from the identification

$$H^n(A,\mathbb{A}_f)\cong \mathrm{Hom}(\bigwedge^n\Lambda,\mathbb{A}_f)\cong \mathrm{Hom}(\bigwedge^n(\Lambda\otimes\mathbb{A}_f),\mathbb{A}_f)\cong \mathrm{Hom}(\bigwedge^n V_fA,\mathbb{A}_f)$$

(or, equivalently, from identifying $H^n(A,\mathbb{A}_f)$ with étale cohomology).

THEOREM 14.13. *Let s be a hodge tensor on an abelian variety A over $\mathbb{C}$, and let $s_{\mathbb{A}_f}$ be the image of s the $\mathbb{A}_f$-cohomology. For any automorphism σ of $\mathbb{C}$, there exists a hodge tensor ${}^{\sigma}s$ on σA (necessarily unique) such that $({}^{\sigma}s)_{\mathbb{A}_f}=\sigma(s_{\mathbb{A}_f})$.*

PROOF. This is the main theorem of Deligne 1982. [Interestingly, the theory of locally symmetric varieties is used in the proof.] □

As an alternative to using Deligne's theorem, one can apply the following result (note, however, that the above approach has the advantage of giving a description of the points of the canonical model with coordinates in any field containing the reflex field).

PROPOSITION 14.14. *Let $(G,X)\hookrightarrow(G',X')$ be an inclusion of Shimura data; if* $\mathrm{Sh}(G',X')$ *has canonical model, so also does* $\mathrm{Sh}(G,X)$.

PROOF. This follows easily from 5.16. □

Shimura varieties of abelian type. Deligne (1979, 2.7.10) defines the notion of a canonical model of a *connected* Shimura variety $\mathrm{Sh}^{\circ}(G,X)$. This is an inverse system of connected varieties over $\mathbb{Q}^{\mathrm{al}}$ endowed with the action of a large group (a mixture of a galois group and an adèlic group). A key result is the following.

THEOREM 14.15. *Let (G,X) be a Shimura datum and let X^+ be a connected component of X. Then* $\mathrm{Sh}(G,X)$ *has a canonical model if and only if* $\mathrm{Sh}^{\circ}(G^{\mathrm{der}},X^+)$ *has a canonical model.*

PROOF. See Deligne 1979, 2.7.13. □

Thus, for example, if (G_1,X_1) and (G_2,X_2) are Shimura data such that $(G_1^{\mathrm{der}},X_1^+)\approx(G_2^{\mathrm{der}},X_2^+)$, and one of $\mathrm{Sh}(G_1,X_1)$ or $\mathrm{Sh}(G_2,X_2)$ has a canonical model, then they both do.

The next result is more obvious (ibid. 2.7.11).

PROPOSITION 14.16. *(a) Let* (G_i, X_i) $(1 \leq i \leq m)$ *be connected Shimura data. If each connected Shimura variety* $\mathrm{Sh}^\circ(G_i, X_i)$ *has a canonical model* $M^\circ(G_i, X_i)$*, then* $\prod_i M^\circ(G_i, X_i)$ *is a canonical model for* $\mathrm{Sh}^\circ(\prod_i G_i, \prod_i X_i)$.

(b) Let $(G_1, X_1) \to (G_2, X_2)$ *be an isogeny of connected Shimura data. If* $\mathrm{Sh}^\circ(G_1, X_1)$ *has a canonical model, then so also does* $\mathrm{Sh}^\circ(G_2, X_2)$.

More precisely, in case (b) of the theorem, let $G^{\mathrm{ad}}(\mathbb{Q})_1^+$ and $G^{\mathrm{ad}}(\mathbb{Q})_2^+$ be the completions of $G^{\mathrm{ad}}(\mathbb{Q})^+$ for the topologies defined by the images of congruence subgroups in $G_1(\mathbb{Q})^+$ and $G_2(\mathbb{Q})^+$ respectively; then the canonical model for $\mathrm{Sh}^\circ(G_2, X_2)$ is the quotient of the canonical model for $\mathrm{Sh}^\circ(G_2, X_2)$ by the kernel of $G^{\mathrm{ad}}(\mathbb{Q})_1^+ \to G^{\mathrm{ad}}(\mathbb{Q})_2^+$.

We can now prove the existence of canonical models for all Shimura varieties of abelian type. For a connected Shimura variety of primitive type, the existence follows from (14.15) and the existence of canonical models for Shimura varieties of hodge type (see above). Now (14.16) proves the existence for all connected Shimura varieties of abelian type, and (14.16) proves the existence for all Shimura varieties of abelian type.

REMARK 14.17. The above proof is only an existence proof: it gives little information about the canonical model. For the Shimura varieties it treats, Theorem 9.4 can be used to construct canonical models and give a description of the points of the canonical model in any field containing the reflex field.

General Shimura varieties. There is an approach that proves the existence of canonical models for all Shimura varieties, and is largely independent of that discussed above except that it assumes the existence[18] of canonical models for Shimura varieties of type A_1 (and it uses (14.15) and (14.16)).

The essential idea is the following. Let (G, X) be a connected Shimura datum with G the group over $\mathbb{Q}$ obtained from a simple group H over a totally real field F by restriction of scalars.

Assume first that H splits over a CM-field of degree 2 over F. Then there exist many homomorphisms $H_i \to H$ from groups of type A_1 into H. From this, we get many inclusions

$$\mathrm{Sh}^\circ(G_i, X_i) \hookrightarrow \mathrm{Sh}^\circ(G, X)$$

where G_i is the restriction of scalars of H_i. From this, and the existence of canonical models for the $\mathrm{Sh}^\circ(G_i, X_i)$, it is possible to prove the existence of the canonical model for $\mathrm{Sh}^\circ(G, X)$.

In the general case, there will be a totally real field F' containing F and such that $H_{F'}$ splits over a CM-field of degree 2 over F. Let G_* be the restriction of scalars of $H_{F'}$. Then there is an inclusion $(G, X) \hookrightarrow (G_*, X_*)$ of connected Shimura data, and the existence of a canonical model for $\mathrm{Sh}^\circ(G_*, X_*)$ implies the existence of a canonical model for $\mathrm{Sh}^\circ(G, X)$ (cf. 14.14).

For the details, see Borovoi 1984, 1987 and Milne 1983.

Final remark: rigidity. One might expect that if one modified the condition (54), for example, by replacing $r_x(s)$ with $r_x(s)^{-1}$, then one would arrive at a modified notion of canonical model, and the same theorems would hold. This is not true: the condition (54) is the *only* one for which canonical models can exist.

[18]In fact, the approach assumes a stronger statement for Shimura varieties of type A_1, namely, Langlands's conjugation conjecture, and it proves Langlands's conjecture for all Shimura varieties.

In fact, if G is adjoint, then the Shimura variety $\mathrm{Sh}(G,X)$ has no automorphisms commuting with the action of $G(\mathbb{A}_f)$ (Milne 1983, 2.7), from which it follows that the canonical model is the *only* model of $\mathrm{Sh}(G,X)$ over $E(G,X)$, and we know that for the canonical model the reciprocity law at the special points is given by (54).

NOTES. The concept of a canonical model characterized by reciprocity laws at special points is due to Shimura, and the existence of such models was proved for major families by Shimura, Miyake, and Shih. Shimura recognized that to have a canonical model it is necessary to have a reductive group, but for him the semisimple group was paramount: in our language, given a connected Shimura datum (H,Y), he asked for Shimura datum (G,X) such that $(G^{\mathrm{der}},X^+)=(H,Y)$ and $\mathrm{Sh}(G,X)$ has a canonical model (see his talk at the 1970 International Congress Shimura 1971). In his Bourbaki report on Shimura's work (1971b), Deligne placed the emphasis on reductive groups, thereby enlarging the scope of the field.

15. Abelian varieties over finite fields

For each Shimura datum (G,X), we now have a canonical model $\mathrm{Sh}(G,X)$ of the Shimura variety over its reflex field $E(G,X)$. In order, for example, to understand the zeta function of the Shimura variety or the galois representations occurring in its cohomology, we need to understand the points on the canonical model when we reduce it modulo a prime of $E(G,X)$. After everything we have discussed, it would be natural to do this in terms of abelian varieties (or motives) over the finite field plus additional structure. However, such a description will not be immediately useful — what we want is something more combinatorial, which can be plugged into the trace formula. The idea of Langlands and Rapoport (1987) is to give an elementary definition of a category of "fake" abelian varieties (better, abelian motives) over the algebraic closure of a finite field that looks just like the true category, and to describe the points in terms of it. In this section, I explain how to define such a category.

Semisimple categories. An object of an abelian category M is ***simple*** if it has no proper nonzero subobjects. Let F be a field. By an ***F-category***, I mean an additive category in which the Hom-sets $\mathrm{Hom}(x,y)$ are finite dimensional F-vector spaces and composition is F-bilinear. An F-category M is said to be ***semisimple*** if it is abelian and every object is a direct sum (necessarily finite) of simple objects.

If e is simple, then a nonzero morphism $e\to e$ is an isomorphism. Therefore, $\mathrm{End}(e)$ is a division algebra over F. Moreover, $\mathrm{End}(re)\cong M_r(\mathrm{End}(e))$. Here re denotes the direct sum of r copies of e. If e' is a second simple object, then either $e\approx e'$ or $\mathrm{Hom}(e,e')=0$. Therefore, if $x=\sum r_ie_i$ $(r_i\geq 0)$ and $y=\sum s_ie_i$ $(s_i\geq 0)$ are two objects of M expressed as sums of copies of simple objects e_i with $e_i\not\approx e_j$ for $i\neq j$, then

$$\mathrm{Hom}(x,y)=\prod M_{s_i,r_i}(\mathrm{End}(e_i)).$$

Thus, the category M is described up to equivalence by:

(a) the set $\Sigma(\mathsf{M})$ of isomorphism classes of simple objects in M;
(b) for each $\sigma\in\Sigma$, the isomorphism class $[D_\sigma]$ of the endomorphism algebra D_σ of a representative of σ.

We call $(\Sigma(\mathsf{M}),([D_\sigma])_{\sigma\in\Sigma(\mathsf{M})})$ the ***numerical invariants*** of M.

Division algebras; the Brauer group. We shall need to understand what the set of isomorphism classes of division algebras over a field F look like.

Recall the definitions: by an F-***algebra***, we mean a ring A containing F in its centre and finite dimensional as F-vector space; if F equals the centre of A, then A is called a ***central*** F-algebra; a ***division algebra*** is an algebra in which every nonzero element has an inverse; an F-algebra A is ***simple*** if it contains no two-sided ideals other than 0 and A. By a theorem of Wedderburn, the simple F-algebras are the matrix algebras over division F-algebras.

EXAMPLE 15.1. (a) If F is algebraically closed or finite, then every central division algebra is isomorphic to F.

(b) Every central division algebra over $\mathbb{R}$ is isomorphic either to $\mathbb{R}$ or to the (usual) quaternion algebra:

$$\mathbb{H} = \mathbb{C} \oplus \mathbb{C}j, \quad j^2 = -1, \quad jzj^{-1} = \overline{z} \quad (z \in \mathbb{C}).$$

(c) Let F be a p-adic field (finite extension of $\mathbb{Q}_p$), and let π be a prime element of $\mathcal{O}_F$. Let L be an unramified extension field of F of degree n, and let σ denote the Frobenius generator of $\mathrm{Gal}(L/F)$ — σ acts as $x \mapsto x^p$ on the residue field. For each i, $1 \leq i \leq n$, define

$$D_{i,n} = L \oplus La \oplus \cdots \oplus La^{n-1}, \quad a^n = \pi^i, \quad aza^{-1} = \sigma(z) \quad (z \in L).$$

Then $D_{i,n}$ is a central simple algebra over F, which is a division algebra if and only if $\gcd(i,n) = 1$. Every central division algebra over F is isomorphic to $D_{i,n}$ for exactly one relatively prime pair (i,n) (CFT, IV 4.2).

If B and B' are central simple F-algebras, then so also is $B \otimes_F B'$ (CFT, 2.8). If D and D' are central division algebras, then Wedderburn's theorem shows that $D \otimes_F D' \approx M_r(D'')$ for some r and some central division algebra D'' well-defined up to isomorphism, and so we can set

$$[D][D'] = [D''].$$

This law of composition is obviously, and $[F]$ is an identity element. Let D^{opp} denote the opposite algebra to D (the same algebra but with the multiplication reversed: $a^{\mathrm{opp}}b^{\mathrm{opp}} = (ba)^{\mathrm{opp}}$). Then (CFT, IV 2.9)

$$D \otimes_F D^{\mathrm{opp}} \cong \mathrm{End}_{F\text{-linear}}(D) \approx M_r(F),$$

and so $[D][D^{\mathrm{opp}}] = [F]$. Therefore, the isomorphism classes of central division algebras over F (equivalently, the isomorphism classes of central simple algebras over F) form a group, called the ***Brauer group*** of F.

EXAMPLE 15.2. (a) The Brauer group of an algebraically closed field or a finite field is zero.

(b) The Brauer group $\mathbb{R}$ has order two: $\mathrm{Br}(\mathbb{R}) \cong \frac{1}{2}\mathbb{Z}/\mathbb{Z}$.

(c) For a p-adic field F, the map $[D_{n,i}] \mapsto \frac{i}{n} \mod \mathbb{Z}$ is an isomorphism $\mathrm{Br}(F) \cong \mathbb{Q}/\mathbb{Z}$.

(d) For a number field F and a prime v, write inv_v for the canonical homomorphism $\mathrm{Br}(F_v) \to \mathbb{Q}/\mathbb{Z}$ given by (a,b,c) (so inv_v is an isomorphism except when v is real or complex, in which case it has image $\frac{1}{2}\mathbb{Z}/\mathbb{Z}$ or 0). For a central simple algebra B over F, $[B \otimes_F F_v] = 0$ for almost all v, and the sequence

$$0 \longrightarrow \mathrm{Br}(F) \xrightarrow{[B]\mapsto[B\otimes_F F_v]} \oplus \mathrm{Br}(F_v) \xrightarrow{\sum \mathrm{inv}_v} \mathbb{Q}/\mathbb{Z} \longrightarrow 0.$$

is exact.

Statement (d) is shown in the course of proving the main theorem of class field theory by the cohomological approach (CFT, VIII 2.2). It says that to give a division algebra over F (up to isomorphism) is the same as to give a family $(i_v) \in \bigoplus_{v \text{ finite}} \mathbb{Q}/\mathbb{Z} \oplus \bigoplus_{v \text{ real}} \frac{1}{2}\mathbb{Z}/\mathbb{Z}$ such that $\sum i_v = 0$.

The key tool in computing Brauer groups is an isomorphism

$$\mathrm{Br}(F) \cong H^2(F, \mathbb{G}_m) \overset{\text{df}}{=} H^2(\mathrm{Gal}(F^{\mathrm{al}}/F), F^{\mathrm{al}\times}) \overset{\text{df}}{=} \varinjlim H^2(\mathrm{Gal}(L/F), L^\times).$$

The last limit is over the fields $L \subset F^{\mathrm{al}}$ of finite degree and galois over $\mathbb{Q}$. This isomorphism can be most elegantly defined as follows. Let D be a central simple division of degree n^2 over F, and assume that D contains a subfield L of degree n over F and galois over F. Then each $\beta \in D$ normalizing L defines an element $x \mapsto \beta x \beta^{-1}$ of $\mathrm{Gal}(L/F)$, and the Noether-Skolem theorem (CFT, IV 2.10) shows that every element of $\mathrm{Gal}(L/F)$ arises in this way. Because L is its own centralizer (ibid., 3.4), the sequence

$$1 \to L^\times \to N(L) \to \mathrm{Gal}(L/F) \to 1$$

is exact. For each $\sigma \in \mathrm{Gal}(L/F)$, choose an $s_\sigma \in N(L)$ mapping to σ, and let

$$s_\sigma \cdot s_\tau = d_{\sigma,\tau} \cdot s_{\sigma\tau}, \quad d_{\sigma,\tau} \in L^\times.$$

Then $(d_{\sigma,\tau})$ is a 2-cocycle whose cohomology class is independent of the choice of the family (s_σ). Its class in $H^2(\mathrm{Gal}(L/F), L^\times) \subset H^2(F, \mathbb{G}_m)$ is the cohomology class of $[D]$.

EXAMPLE 15.3. Let L be the completion of $\mathbb{Q}_p^{\mathrm{un}}$ (equal to the field of fractions of the ring of Witt vectors with coefficients in $\mathbb{F}$), and let σ be the automorphism of L inducing $x \mapsto x^p$ on its residue field. An ***isocrystal*** is a finite dimensional L-vector space V equipped with a σ-linear isomorphism $F\colon V \to V$. The category Isoc of isocrystals is a semisimple $\mathbb{Q}_p$-linear category with $\Sigma(\mathsf{Isoc}) = \mathbb{Q}$, and the endomorphism algebra of a representative of the isomorphism class λ is a division algebra over $\mathbb{Q}_p$ with invariant λ. If $\lambda \geq 0$, $\lambda = r/s$, $\gcd(r,s) = 1$, $s > 0$, then E^λ can be taken to be $(\mathbb{Q}_p/(T^r - p^s)) \otimes_{\mathbb{Q}_p} L$, and if $\lambda < 0$, E^λ can be taken to be the dual of $E^{-\lambda}$. See Demazure 1972, Chap. IV.

Abelian varieties. Recall (p334) that $\mathsf{AV}^0(k)$ is the category whose objects are the abelian varieties over k, but whose homs are $\mathrm{Hom}^0(A,B) = \mathrm{Hom}(A,B) \otimes \mathbb{Q}$. It follows from results of Weil that $\mathsf{AV}^0(k)$ is a semisimple $\mathbb{Q}$-category with the simple abelian varieties (see p334) as its simple objects. Amazingly, when k is finite, we know its numerical invariants.

Abelian varieties over $\mathbb{F}_q$, $q = p^n$. Recall that a Weil q-integer is an algebraic integer such that, for every embedding $\rho\colon \mathbb{Q}[\pi] \to \mathbb{C}$, $|\rho\pi| = q^{\frac{1}{2}}$. Two Weil q-integers π and π' are ***conjugate*** if there exists an isomorphism $\mathbb{Q}[\pi] \to \mathbb{Q}[\pi']$ sending π to π'.

THEOREM 15.4 (Honda-Tate). *The map* $A \mapsto \pi_A$ *defines a bijection from* $\Sigma(\mathsf{AV}(\mathbb{F}_q))$ *to the set of conjugacy classes of Weil* q*-integers. For any simple* A*, the centre of* $D =_{df} \operatorname{End}^0(A)$ *is* $F = \mathbb{Q}[\pi_A]$*, and for a prime* v *of* F*,*

$$\operatorname{inv}_v(D) = \begin{cases} \frac{1}{2} & \text{if } v \text{ is real} \\ \frac{\operatorname{ord}_v(\pi_A)}{\operatorname{ord}_v(q)}[F_v : \mathbb{Q}_p] & \text{if } v|p \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $2\dim A = [D\colon F]^{\frac{1}{2}} \cdot [F : \mathbb{Q}]$.

In fact, $\mathbb{Q}[\pi]$ can only have a real prime if $\pi = \sqrt{p^n}$. Let $W_1(q)$ be the set of Weil q-integers in $\mathbb{Q}^{\mathrm{al}} \subset \mathbb{C}$. Then the theorem gives a bijection

$$\Sigma(\mathsf{AV}^0(\mathbb{F}_q)) \to \Gamma\backslash W_1(q), \quad \Gamma = \operatorname{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q}).$$

NOTES. Except for the statement that every π_A arises from an A, the theorem is due to Tate. That every Weil q-integer arises from an abelian variety was proved (using 10.10) by Honda. See Tate 1969 for a discussion of the theorem.

Abelian varieties over $\mathbb{F}$. We shall need a similar result for an algebraic closure $\mathbb{F}$ of $\mathbb{F}_p$.

If π is a Weil p^n-integer, then π^m is a Weil p^{mn}-integer, and so we have a homomorphism $\pi \mapsto \pi^m : W_1(p^n) \to W_1(p^{nm})$. Define

$$W_1 = \varinjlim W_1(p^n).$$

If $\pi \in W_1$ is represented by $\pi_n \in W_1(p^n)$, then $\pi_n^m \in W_1(p^{nm})$ also represents π, and $\mathbb{Q}[\pi_n] \supset \mathbb{Q}[\pi_n^m]$. Define $\mathbb{Q}\{\pi\}$ to be the field of smallest degree over $\mathbb{Q}$ generated by a representative of π.

Every abelian variety over $\mathbb{F}$ has a model defined over a finite field, and if two abelian varieties over a finite field become isomorphic over $\mathbb{F}$, then they are isomorphic already over a finite field. Let A be an abelian variety over $\mathbb{F}_q$. When we regard A as an abelian variety over $\mathbb{F}_{q^m}$, then the Frobenius map is raised to the m^{th}-power (obviously): $\pi_{A_{\mathbb{F}_{q^m}}} = \pi_A^m$.

Let A be an abelian variety defined over $\mathbb{F}$, and let A_0 be a model of A over $\mathbb{F}_q$. The above remarks show that $s_A(v) =_{\mathrm{df}} \frac{\operatorname{ord}_v(\pi_{A_0})}{\operatorname{ord}_v(q)}$ is independent of the choice of A_0. Moreover, for any $\rho\colon \mathbb{Q}[\pi_{A_0}] \hookrightarrow \mathbb{Q}^{\mathrm{al}}$, the Γ-orbit of the element π_A of W_1 represented by $\rho\pi_{A_0}$ depends only on A.

THEOREM 15.5. *The map* $A \mapsto \Gamma\pi_A$ *defines a bijection* $\Sigma(\mathsf{AV}^0(\mathbb{F})) \to \Gamma\backslash W_1$*. For any simple* A*, the centre of* $D =_{df} \operatorname{End}^0(A)$ *is isomorphic to* $F = \mathbb{Q}\{\pi_A\}$*, and for any prime* v *of* F*,*

$$\operatorname{inv}_v(D) = \begin{cases} \frac{1}{2} & \text{if } v \text{ is real} \\ s_A(v) \cdot [F_v : \mathbb{Q}_p] & \text{if } v|p \\ 0 & \text{otherwise.} \end{cases}$$

PROOF. This follows from the Honda-Tate theorem and the above discussion. □

Our goal in the remainder of this section is to give an elementary construction of a semisimple $\mathbb{Q}$-category that contains, in a natural way, a category of "fake abelian varieties over $\mathbb{F}$" with the same numerical invariants as $\mathsf{AV}^0(\mathbb{F})$.

For the remainder of this section F is a field of characteristic zero.

Tori and their representations. Let T be a torus over F split by a galois extension L/F with galois group Γ. As we noted on p276, to give a representation ρ of T on an F-vector space V amounts to giving an $X^*(T)$-grading $V(L) = \bigoplus_{\chi \in X^*(T)} V_\chi$ of $V(L)$ with the property that $\sigma V_\chi = V_{\sigma\chi}$ for all $\sigma \in \Gamma$ and $\chi \in X^*(T)$. In this, L/F can be an infinite galois extension.

PROPOSITION 15.6. *Let $\Gamma = \mathrm{Gal}(F^{\mathrm{al}}/F)$. The category of representations $\mathsf{Rep}(T)$ of T on F-vector spaces is semisimple. The set of isomorphism classes of simple objects is in natural one-to-one correspondence with the orbits of Γ acting on $X^*(T)$, i.e., $\Sigma(\mathsf{Rep}(T)) = \Gamma \backslash X^*(T)$. If $V_{\Gamma\chi}$ is a simple object corresponding to $\Gamma\chi$, then $\dim(V_{\Gamma\chi})$ is the order of $\Gamma\chi$, and*

$$\mathrm{End}(V_\chi) \approx F(\chi)$$

where $F(\chi)$ is the fixed field of the subgroup $\Gamma(\chi)$ of Γ fixing χ.

PROOF. Follows easily from the preceding discussion. □

REMARK 15.7. Let $\chi \in X^*(T)$, and let $\Gamma(\chi)$ and $F(\chi)$ be as in the proposition. Then $\mathrm{Hom}(F(\chi), F^{\mathrm{al}}) \cong \Gamma/\Gamma(\chi)$, and so $X^*((\mathbb{G}_m)_{F(\chi)/F}) = \mathbb{Z}^{\Gamma/\Gamma(\chi)}$. The map

$$\textstyle\sum n_\sigma \sigma \mapsto \sum n_\sigma \sigma\chi \colon \mathbb{Z}^{\Gamma/\Gamma(\chi)} \to X^*(T)$$

defines a homorphism

$$T \to (\mathbb{G}_m)_{F(\chi)/F}. \tag{58}$$

From this, we get a homomorphism of cohomology groups

$$H^2(F,T) \to H^2(F,(\mathbb{G}_m)_{F(\chi)/F}).$$

But Shapiro's lemma (CFT, II 1.11) shows that $H^2(F,(\mathbb{G}_m)_{F(\chi)/F}) \cong H^2(F(\chi), \mathbb{G}_m)$, which is the Brauer group of $F(\chi)$. On composing these maps, we get a homomorphism

$$H^2(F,T) \to \mathrm{Br}(F(\chi)). \tag{59}$$

The proposition gives a natural construction of a semisimple category M with $\Sigma(\mathsf{M}) = \Gamma\backslash N$, where N is any finitely generated $\mathbb{Z}$-module equipped with a continuous action of Γ. However, the simple objects have commutative endomorphism algebras. To go further, we need to look at new type of structure.

Affine extensions. Let L/F be a Galois extension of fields with Galois group Γ, and let G be an algebraic group over F. In the following, we consider only extensions

$$1 \to G(L) \to E \to \Gamma \to 1$$

in which the action of Γ on $G(L)$ defined by the extension is the natural action, i.e.,

$$\text{if } e_\sigma \mapsto \sigma, \text{ then } e_\sigma g e_\sigma^{-1} = \sigma g \quad (e_\sigma \in E,\ \sigma \in \Gamma,\ g \in T(F^{\mathrm{al}})).$$

For example, there is always the ***split extension*** $E_G =_{\mathrm{df}} G(L) \rtimes \Gamma$.

An extension E is ***affine*** if its pull-back to some open subgroup of Γ is split. Equivalently, if for the σ in some open subgroup of Γ, there exist $e_\sigma \mapsto \sigma$ such that $e_{\sigma\tau} = e_\sigma e_\tau$. We sometimes call such an E an L/F-affine extension with ***kernel*** G.

Consider an extension

$$1 \to T \to E \to \Gamma \to 1$$

with T commutative. If E is affine, then it is possible to choose the e_σ's so that the 2-cocycle $d\colon \Gamma\times\Gamma\to T(L)$ defined by

$$e_\sigma e_\tau = d_{\sigma,\tau} e_\sigma e_\tau, \quad d_{\sigma,\tau}\in T(F^{\mathrm{al}}).$$

is continuous. Thus, in this case E defines a class $cl(E)\in H^2(F,T)$.

A ***homomorphism*** of affine extensions is a commutative diagram

$$\begin{array}{ccccccccc}
1 & \longrightarrow & G_1(L) & \longrightarrow & E_1 & \longrightarrow & \Gamma & \longrightarrow & 1\\
 & & \downarrow & & \downarrow{\scriptstyle\phi} & & \| & & \\
1 & \longrightarrow & G_2(L) & \longrightarrow & E_2 & \longrightarrow & \Gamma & \longrightarrow & 1
\end{array}$$

such that the restriction of the homomorphism ϕ to $G_1(L)$ is defined by a homomorphism of algebraic groups (over L). A ***morphism*** $\phi\to\phi'$ of homomorphisms $E_1\to E_2$ is an element of g of $G_2(L)$ such that $\mathrm{ad}(g)\circ\phi=\phi'$, i.e., such that

$$g\cdot\phi(e)\cdot g^{-1}=\phi'(e), \quad \text{all } e\in E_1.$$

For a vector space V over F, let E_V be the split affine extension defined by the algebraic group $\mathrm{GL}(V)$. A ***representation*** of an affine extension E is a homomorphism $E\to E_V$.

Remark 15.8. To give a representation of E_G on E_V is the same as to give a representation of G on V. More precisely, the functor $\mathsf{Rep}(G)\to\mathsf{Rep}(E_G)$ is an equivalence of categories. The proof of this uses that $H^1(F,\mathrm{GL}(V))=1$.

Proposition 15.9. *Let E be an L/F-affine extension whose kernel is a torus T split by L. The category $\mathsf{Rep}(E)$ is a semisimple F-category with $\Sigma(\mathsf{Rep}(E))=\Gamma\backslash X^*(T)$. Let $V_{\Gamma\chi}$ be a simple representation of E corresponding to $\Gamma\chi\in\Gamma\backslash X^*(T)$. Then, $D=\mathrm{End}(V_{\Gamma_\chi})$ has centre $F(\chi)$, and its class in $\mathrm{Br}(F(\chi))$ is the image of $cl(E)$ under the homomorphism (59).*

Proof. Omitted (but it is not difficult). □

We shall also need to consider affine extensions in which the kernel is allowed to be a protorus, i.e., the limit of an inverse system of tori. For $T=\varprojlim T_i$, $X^*(T)=\varinjlim X^*(T_i)$, and $T\mapsto X^*(T)$ defines an equivalence from the category of protori to the category of free $\mathbb{Z}$-modules with a continuous action of Γ. Here continuous means that every element of the module is fixed by an open subgroup of Γ. Let $L=F^{\mathrm{al}}$. By an ***affine extension with kernel*** T, we mean an exact sequence

$$1\to T(F^{\mathrm{al}})\to E\to\Gamma\to 1$$

whose push-out

$$1\to T_i(F^{\mathrm{al}})\to E_i\to\Gamma\to 1$$

by $T(F^{\mathrm{al}})\to T_i(F^{\mathrm{al}})$ is an affine extension in the previous sense. A representation of such an extension is defined exactly as before.

Remark 15.10. Let

$$\begin{array}{ccc}
L & \overset{\subset}{\text{———}} & L'\\
\Big|{\scriptstyle\Gamma} & & \Big|{\scriptstyle\Gamma'}\\
F & \overset{\subset}{\text{———}} & F'
\end{array}$$

be a diagram of fields in which L'/F' is Galois with group Γ'. From an L/F-affine extension

$$1 \to G(L) \to E \to \Gamma \to 1$$

with kernel G we obtain an L'/F'-affine extension

$$1 \to G(L') \to E' \to \Gamma' \to 1$$

with kernel $G_{F'}$ by pulling back by $\sigma \mapsto \sigma|L\colon \Gamma' \to \Gamma$ and pushing out by $G(L) \to G(L'))$.

EXAMPLE 15.11. Let $\mathbb{Q}_p^{\mathrm{un}}$ be a maximal unramified extension of $\mathbb{Q}_p$, and let L_n be the subfield of $\mathbb{Q}_p^{\mathrm{un}}$ of degree n over $\mathbb{Q}_p$. Let $\Gamma_n = \mathrm{Gal}(L_n/\mathbb{Q}_p)$, let $D_{1,n}$ be the division algebra in (15.1c), and let

$$1 \to L_n^{\times} \to N(L_n^{\times}) \to \Gamma_n \to 1$$

be the corresponding extension. Here $N(L_n^{\times})$ is the normalizer of $L_n^{\times}$ in $D_{1,n}$:

$$N(L_n^{\times}) = \bigsqcup\nolimits_{0 \le i \le n-1} L_n^{\times} a^i.$$

This is an $L_n/\mathbb{Q}_p$-affine extension with kernel $\mathbb{G}_m$. On pulling back by $\Gamma \to \Gamma_n$ and pushing out by $L_n^{\times} \to \mathbb{Q}_p^{\mathrm{un}\times}$, we obtain a $\mathbb{Q}_p^{\mathrm{un}\times}/\mathbb{Q}_p$-affine extension D_n with kernel $\mathbb{G}_m$. From a representation of D_n we obtain a vector space V over $\mathbb{Q}_p^{\mathrm{un}}$ equipped with a σ-linear map F (the image of $(1,a)$ is (F,σ)). On tensoring this with the completion L of $\mathbb{Q}_p^{\mathrm{un}}$, we obtain an isocrystal that can be expressed as a sum of E^{λ}'s with $\lambda \in \frac{1}{n}\mathbb{Z}$.

Note that there is a canonical section to $N(L_n^{\times}) \to \Gamma_n$, namely, $\sigma^i \mapsto a^i$, which defines a canonical section to $D_n \to \Gamma$.

There is a homomorphism $D_{nm} \to D_n$ whose restriction to the kernel is multiplication by m. The inverse limit of this system is a $\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p$-affine extension D with kernel $\mathbb{G} =_{\mathrm{df}} \varprojlim \mathbb{G}_m$. Note that $X^*(\mathbb{G}) = \varinjlim \frac{1}{n}\mathbb{Z}/\mathbb{Z} = \mathbb{Q}$. There is a natural functor from $\mathsf{Rep}(D)$ to the category of isocrystals, which is faithful and essentially surjective on objects but not full. We call D the ***Dieudonné affine extension***.

The affine extension $\mathfrak{P}$. Let $W(p^n)$ be the subgroup of $\mathbb{Q}^{\mathrm{al}\times}$ generated by $W_1(p^n)$, and let $W = \varinjlim W(p^n)$. Then W is a free $\mathbb{Z}$-module of infinite rank with a continuous action of $\Gamma = \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$. For $\pi \in W$, we define $\mathbb{Q}\{\pi\}$ to be the smallest field generated by a representative of π. If π is represented by $\pi_n \in W(p^n)$ and $|\rho(\pi_n)| = (p^n)^{m/2}$, we say that π has weight m and we write

$$s_\pi(v) = \frac{\mathrm{ord}_v(\pi_n)}{\mathrm{ord}_v(q)}.$$

THEOREM 15.12. *Let P be the protorus over $\mathbb{Q}$ with $X^*(P) = W$. Then there exists an affine extension*

$$1 \to P(\mathbb{Q}^{\mathrm{al}}) \to \mathfrak{P} \to \Gamma \to 1$$

such that

(a) $\Sigma(\mathsf{Rep}(\mathfrak{P})) = \Gamma\backslash W$;
(b) *for $\pi \in W$, let $D(\pi) = \mathrm{End}(V_{\Gamma\pi})$ where $V_{\Gamma\pi}$ is a representation corresponding to $\Gamma\pi$; then $D(\pi)$ is isomorphic to the division algebra D with*

centre $\mathbb{Q}\{\pi\}$ *and the invariants*

$$\mathrm{inv}_v(D) = \begin{cases} (\frac{1}{2})^{wt(\pi)} & \text{if } v \text{ is real} \\ s_\pi(v)\cdot[\mathbb{Q}\{\pi\}_v : \mathbb{Q}_p] & \text{if } v|p \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $\mathfrak{P}$ *is unique up to isomorphism.*

PROOF. Let $c(\pi)$ denote the class in $\mathrm{Br}(\mathbb{Q}\{\pi\})$ of the division algebra D in (b). To prove the result, we have to show that there exists a unique class in $H^2(\mathbb{Q}, P)$ mapping to $c(\pi)$ in $\mathrm{Br}(\mathbb{Q}\{\pi\})$ for all π:

$$c \mapsto (c(\pi))\colon H^2(\mathbb{Q}, P) \overset{(59)}{\to} \textstyle\prod_{\Gamma\pi\in\Gamma\backslash W} \mathrm{Br}(\mathbb{Q}\{\pi\}).$$

This is an exercise in galois cohomology, which, happily, is easier than it looks. □

We call a representation of $\mathfrak{P}$ a ***fake motive*** over $\mathbb{F}$, and a ***fake abelian variety*** if its simple summands correspond to $\pi \in \Gamma\backslash W_1$. Note that the category of fake abelian varieties is a semisimple $\mathbb{Q}$-category with the same numerical invariants as $\mathsf{AV}^0(\mathbb{F})$.

The local form $\mathfrak{P}_l$ *of* $\mathfrak{P}$. Let l be a prime of $\mathbb{Q}$, and choose a prime w_l of $\mathbb{Q}^{\mathrm{al}}$ dividing l. Let $\mathbb{Q}_l^{\mathrm{al}}$ be the algebraic closure of $\mathbb{Q}_l$ in the completion of $\mathbb{Q}^{\mathrm{al}}$ at w_l. Then $\Gamma_l =_{\mathrm{df}} \mathrm{Gal}(\mathbb{Q}_l^{\mathrm{al}}/\mathbb{Q}_l)$ is a closed subgroup of $\Gamma =_{\mathrm{df}} \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$, and we have a diagram

(60)
$$\begin{tikzcd} \mathbb{Q}^{\mathrm{al}} \arrow[r, dash] \arrow[d, dash, "\Gamma"] & \mathbb{Q}_l^{\mathrm{al}} \arrow[d, dash, "\Gamma_l"] \\ \mathbb{Q} \arrow[r, dash] & \mathbb{Q}_l. \end{tikzcd}$$

From $\mathfrak{P}$ we obtain a $\mathbb{Q}_l^{\mathrm{al}}/\mathbb{Q}_l$-affine extension $\mathfrak{P}(l)$ by pulling back by $\Gamma_l \to \Gamma$ and pushing out by $P(\mathbb{Q}^{\mathrm{al}}) \to P(\mathbb{Q}_l^{\mathrm{al}})$ (cf. 15.10).

The $\mathbb{Q}_\ell$*-space attached to a fake motive.* Let $\ell \neq p, \infty$ be a prime of $\mathbb{Q}$.

PROPOSITION 15.13. *There exists a continuous homomorphism* ζ_ℓ *making*

$$\begin{tikzcd} & & & \Gamma_\ell \arrow[dl, "\zeta_\ell"'] \arrow[d, equal] & \\ 1 \arrow[r] & P(\mathbb{Q}_\ell^{\mathrm{al}}) \arrow[r] & \mathfrak{P}(\ell) \arrow[r] & \Gamma_\ell \arrow[r] & 1 \end{tikzcd}$$

commute.

PROOF. To prove this, we have to show that the cohomology class of $\mathfrak{P}$ in $H^2(\mathbb{Q}, P)$ maps to zero in $H^2(\mathbb{Q}_\ell, P)$, but this is not difficult. □

Fix a homomorphism $\zeta_\ell\colon \Gamma_\ell \to \mathfrak{P}(\ell)$ as in the diagram. Let $\rho\colon \mathfrak{P} \to E_V$ be a fake motive. From ρ, we get a homomorphism

$$\rho(\ell)\colon \mathfrak{P}(\ell) \to \mathrm{GL}(V(\mathbb{Q}_\ell^{\mathrm{al}})) \rtimes \Gamma_\ell.$$

For $\sigma \in \Gamma_\ell$, let $(\rho(\ell)\circ\zeta_\ell)(\sigma) = (e_\sigma, \sigma)$. Because ζ_ℓ is a homomorphism, the automorphisms e_σ of $V(\mathbb{Q}_\ell^{\mathrm{al}})$ satisfy

$$e_\sigma \circ \sigma e_\tau = e_{\sigma\tau}, \quad \sigma, \tau \in \Gamma_\ell,$$

and so

$$\sigma \cdot v = e_\sigma(\sigma v)$$

is an action of Γ_ℓ on $V(\mathbb{Q}_\ell^{\mathrm{al}})$, which one can check to be continuous. Therefore (AG 16.14), $V_\ell(\rho) =_{\mathrm{df}} V(\mathbb{Q}_\ell^{\mathrm{al}})^{\Gamma_\ell}$ is a $\mathbb{Q}_\ell$-structure on $V(\mathbb{Q}_\ell^{\mathrm{al}})$. In this way, we get a functor $\rho \mapsto V_\ell(\rho)$ from the category of fake motives over $\mathbb{F}$ to vector spaces over $\mathbb{Q}_\ell$.

The ζ_ℓ can be chosen in such a way that the spaces $V_\ell(\rho)$ contain lattices $\Lambda_\ell(\rho)$ that are well-defined for almost all $\ell \neq p$, which makes it possible to define

$$V_f^p(\rho) = \prod\nolimits_{\ell \neq p,\infty} (V_\ell(\rho) \colon \Lambda_\ell(\rho)).$$

It is a free module over $\mathbb{A}_f^p =_{\mathrm{df}} \prod_{\ell \neq p,\infty} (\mathbb{Q}_\ell \colon \mathbb{Z}_\ell)$.

The isocrystal of a fake motive. Choose a prime w_p of $\mathbb{Q}^{\mathrm{al}}$ dividing p, and let $\mathbb{Q}_p^{\mathrm{un}}$ and $\mathbb{Q}_p^{\mathrm{al}}$ denote the subfields of the completion of $\mathbb{Q}^{\mathrm{al}}$ at w_p. Then $\Gamma_p =_{\mathrm{df}} \mathrm{Gal}(\mathbb{Q}_p^{\mathrm{al}}/\mathbb{Q}_p)$ is a closed subgroup of $\Gamma =_{\mathrm{df}} \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ and $\Gamma_p^{\mathrm{un}} =_{\mathrm{df}} \mathrm{Gal}(\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p)$ is a quotient of Γ_p.

PROPOSITION 15.14. *(a) The affine extension $\mathfrak{P}(p)$ arises by pull-back and push-out from a $\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p$-affine extension $\mathfrak{P}(p)^{\mathrm{un}}$.*

(b) There is a homomorphism of $\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p$-extensions $D \to \mathfrak{P}(p)^{\mathrm{un}}$ whose restriction to the kernels, $\mathbb{G} \to P_{\mathbb{Q}_p}$, corresponds to the map on characters $\pi \mapsto s_\pi(w_p) \colon W \to \mathbb{Q}$.

PROOF. (a) This follows from the fact that the image of the cohomology class of $\mathfrak{P}$ in $H^2(\Gamma_p, P(\mathbb{Q}_p^{\mathrm{al}}))$ arises from a cohomology class in $H^2(\Gamma_p^{\mathrm{un}}, P(\mathbb{Q}_p^{\mathrm{un}}))$.

(b) This follows from the fact that the homomorphism $H^2(\mathbb{Q}_p, \mathbb{G}) \to H^2(\mathbb{Q}_p, P_{\mathbb{Q}_p})$ sends the cohomology class of D to that of $\mathfrak{P}(p)^{\mathrm{un}}$. $\square$

In summary:

$$\begin{tikzcd}
1 \arrow[r] & \mathbb{G}_m(\mathbb{Q}_p^{\mathrm{un}}) \arrow[r] \arrow[d] & D \arrow[r] \arrow[d] & \Gamma_p^{\mathrm{un}} \arrow[r] \arrow[d, equal] & 1 \\
1 \arrow[r] & P(\mathbb{Q}_p^{\mathrm{un}}) \arrow[r] \arrow[d] & \mathfrak{P}(p)^{\mathrm{un}} \arrow[r] \arrow[d, phantom, "\vdots"] & \Gamma_p^{\mathrm{un}} \arrow[r] & 1 \\
1 \arrow[r] & P(\mathbb{Q}_p^{\mathrm{al}}) \arrow[r] & \mathfrak{P}(p) \arrow[r] & \Gamma_p \arrow[r] \arrow[u] & 1
\end{tikzcd}$$

A fake motive $\rho \colon \mathfrak{P} \to E_V$ gives rise to a representation of $\mathfrak{P}(p)$, which arises from a representation of $\mathfrak{P}(p)^{\mathrm{un}}$ (cf. the argument in the preceding subsubsection). On composing this with the homomorphism $D \to \mathfrak{P}(p)^{\mathrm{un}}$, we obtain a representation of D, which gives rise to an isocrystal $D(\rho)$ as in (15.11).

Abelian varieties of CM-type and fake abelian varieties. We saw in (10.5) that an abelian variety of CM-type over $\mathbb{Q}^{\mathrm{al}}$ defines an abelian variety over $\mathbb{F}$. Does it also define a fake abelian variety? The answer is yes.

PROPOSITION 15.15. *Let T be a torus over $\mathbb{Q}$ split by a CM-field, and let μ be a cocharacter of T such that $\mu + \iota\mu$ is defined over $\mathbb{Q}$ (here ι is complex conjugation). Then there is a homomorphism, well defined up to isomorphism,*

$$\phi_\mu \colon \mathfrak{P} \to E_T.$$

PROOF. Omitted. $\square$

Let A be an abelian variety of CM-type (E,Φ) over $\mathbb{Q}^{\mathrm{al}}$, and let $T=(\mathbb{G}_m)_{E/\mathbb{Q}}$. Then Φ defines a cocharacter μ_Φ of T (see 12.4(b)), which obviously satisfies the conditions of the proposition. Hence we obtain a homomorphism $\phi\colon \mathfrak{P}\to E_T$. Let $V=H_1(A,\mathbb{Q})$. From ϕ and the representation ρ of T on V we obtain a fake abelian variety $\rho\circ\phi$ such that $V_\ell(\rho\circ\phi)=H_1(A,\mathbb{Q}_\ell)$ (obvious) and $D(\rho)$ is isomorphic to the Dieudonné module of the reduction of A (restatement of the Shimura-Taniyama formula).

Aside 15.16. The category of fake abelian varieties has similar properties to $\mathsf{AV}^0(\mathbb{F})$. By using the $\mathbb{Q}_\ell$-spaces and the isocrystals attached to a fake abelian variety, it is possible to define a $\mathbb{Z}$-linear category with properties similar to $\mathsf{AV}(\mathbb{F})$.

Notes. The affine extension $\mathfrak{P}$ is defined in Langlands and Rapoport 1987, §§1–3, where it is called "die pseudomotivische Galoisgruppe". There an affine extension is called a Galoisgerbe although, rather than a gerbe, it can more accurately be described as a concrete realizations of a groupoid. See also Milne 1992. In the above, I have ignored uniqueness questions, which can be difficult (see Milne 2003).

16. The good reduction of Shimura varieties

We now write $\mathrm{Sh}_K(G,X)$, or just Sh_K, for the canonical model of the Shimura variety over its reflex field.

The points of the Shimura variety with coordinates in the algebraic closure of the rational numbers. When we have a description of the points of the Shimura variety over $\mathbb{C}$ in terms of abelian varieties or motives plus additional data, then the same description holds over $\mathbb{Q}^{\mathrm{al}}$. For example, for the Siegel modular variety attached to a symplectic space (V,ψ), $\mathrm{Sh}_K(\mathbb{Q}^{\mathrm{al}})$ classifies the isomorphism classes of triples $(A,s,\eta K)$ in which A is an abelian variety defined over $\mathbb{Q}^{\mathrm{al}}$, s is an element of $\mathrm{NS}(A)\otimes\mathbb{Q}$ containing a $\mathbb{Q}^\times$-multiple of an ample divisor, and η is a K-orbit of isomorphisms $V(\mathbb{A}_f)\to V_f(A)$ sending ψ to an $\mathbb{A}_f^\times$-multiple of the pairing defined by s. Here $\mathrm{NS}(A)$ is the Nèron-Severi group of A (divisor classes modulo algebraic equivalence).

On the other hand, I do not know a description of $\mathrm{Sh}_K(\mathbb{Q}^{\mathrm{al}})$ when, for example, G^{ad} has factors of type E_6 or E_7 or mixed type D. In these cases, the proof of the existence of a canonical model is quite indirect.

The points of the Shimura variety with coordinates in the reflex field. Over $E=E(G,X)$ the following additional problem arises. Let A be an abelian variety over $\mathbb{Q}^{\mathrm{al}}$. Suppose we know that σA is isomorphic to A for all $\sigma\in\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E)$. Does this imply that A is defined over E? Choose an isomorphism $f_\sigma\colon \sigma A\to A$ for each σ fixing E. A necessary condition that the f_σ arise from a model over E is that they satisfy the cocycle condition: $f_\sigma\circ\sigma f_\tau=f_{\sigma\tau}$. Of course, if the cocycle condition fails for one choice of the f_σ's, we can try another, but there is an obstruction to obtaining a cocycle which lies in the cohomology set $H^2(\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E),\mathrm{Aut}(A))$.

Certainly, this obstruction would vanish if $\mathrm{Aut}(A)$ were trivial. One may hope that the automorphism group of an abelian variety (or motive) plus data in the family classified by $\mathrm{Sh}_K(G,X)$ is trivial, at least when K is small. This is so when condition SV5 holds, but not otherwise.

In the Siegel case, the centre of G is $\mathbb{G}_m$ and so SV5 holds. Therefore, provided K is sufficiently small, for any field L containing $E(G,X)$, $\mathrm{Sh}_K(L)$ classifies triples $(A,s,\eta K)$ satisfying the same conditions as when $L=\mathbb{Q}^{\mathrm{al}}$. Here A an abelian variety over L, $s\in \mathrm{NS}(A)\otimes\mathbb{Q}$, and η is an isomorphism $V(\mathbb{A}_f)\to V_f(A)$ such that ηK is stable under the action of $\mathrm{Gal}(L^{\mathrm{al}}/L)$.

In the Hilbert case (4.14), the centre of G is $(\mathbb{G}_m)_{F/\mathbb{Q}}$ for F a totally real field and SV5 fails: $F^\times$ is not discrete in $\mathbb{A}_{F,f}^\times$ because every nonempty open subgroup of $\mathbb{A}_{F,f}^\times$ will contain infinitely many units. In this case, one has a description of $\mathrm{Sh}_K(L)$ when L is algebraically closed, but otherwise all one can say is that $\mathrm{Sh}_K(L)=\mathrm{Sh}_K(L^{\mathrm{al}})^{\mathrm{Gal}(L^{\mathrm{al}}/L)}$.

Hyperspecial subgroups. The modular curve $\Gamma_0(N)\backslash\mathcal{H}_1$ is defined over $\mathbb{Q}$, and it has good reduction at the primes not dividing the level N and bad reduction at the others. Before explaining what is known in general, we need to introduce the notion of a hyperspecial subgroup.

DEFINITION 16.1. Let G be a reductive group over $\mathbb{Q}$ (over $\mathbb{Q}_p$ will do). A subgroup $K\subset G(\mathbb{Q}_p)$ is ***hyperspecial*** if there exists a flat group scheme $\mathcal{G}$ over $\mathbb{Z}_p$ such that

- $\mathcal{G}_{\mathbb{Q}_p}=G$ (i.e., $\mathcal{G}$ extends G to $\mathbb{Z}_p$);
- $\mathcal{G}_{\mathbb{F}_p}$ is a connected reductive group (necessarily of the same dimension as G because of flatness);
- $\mathcal{G}(\mathbb{Z}_p)=K$.

EXAMPLE 16.2. Let $G=\mathrm{GSp}(V,\psi)$. Let Λ be a lattice in $V(\mathbb{Q}_p)$, and let K_p be the stabilizer of Λ. Then K_p is hyperspecial if the restriction of ψ to $\Lambda\times\Lambda$ takes values in $\mathbb{Z}_p$ and is perfect (i.e., induces an isomorphism $\Lambda\to\Lambda^\vee$; equivalently, induces a nondegenerate pairing $\Lambda/p\Lambda\times\Lambda/p\Lambda\to\mathbb{F}_p$). In this case, $\mathcal{G}_{\mathbb{F}_p}$ is again a group of symplectic similitudes over $\mathbb{F}_p$ (at least if $p\neq 2$).

EXAMPLE 16.3. In the PEL-case, in order for there to exist a hyperspecial group, the algebra B must be unramified above p, i.e., $B\otimes_{\mathbb{Q}}\mathbb{Q}_p$ must be a product of matrix algebras over unramified extensions of $\mathbb{Q}_p$. When this condition holds, the description of the hyperspecial groups is similar to that in the Siegel case.

There exists a hyperspecial subgroup in $G(\mathbb{Q}_p)$ if and only if G is ***unramified*** over $\mathbb{Q}_p$, i.e., quasisplit over $\mathbb{Q}_p$ and split over an unramified extension.

For the remainder of this section we fix a hyperspecial subgroup $K_p\subset G(\mathbb{Q}_p)$, and we write $\mathrm{Sh}_p(G,X)$ for the family of varieties $\mathrm{Sh}_{K^p\times K_p}(G,X)$ with K^p running over the compact open subgroups of $G(\mathbb{A}_f^p)$. The group $G(\mathbb{A}_f^p)$ acts on the family $\mathrm{Sh}_p(G,X)$.

The good reduction of Shimura varieties. Roughly speaking, there are two reasons a Shimura variety may have bad reduction at a prime dividing p: the reductive group itself may be ramified at p or p may divide the level. For example, the Shimura curve defined by a quaternion algebra B over $\mathbb{Q}$ will have bad reduction at a prime p dividing the discriminant of B, and (as we noted above) $\Gamma_0(N)\backslash\mathcal{H}_1$ has bad reduction at a prime dividing N. The existence of a hyperspecial subgroup K_p forces G to be unramified at p, and by considering only the varieties $\mathrm{Sh}_{K^pK_p}(G,X)$ we avoid the second problem.

THEOREM 16.4. *Let* $\mathrm{Sh}_p(G,X)$ *be the inverse system of varieties over* $E(G,X)$ *defined by a Shimura datum* (G,X) *of abelian type and a hyperspecial subgroup* $K_p \subset G(\mathbb{Q}_p)$. *Then, except possibly for some small set of primes* p *depending only on* (G,X), $\mathrm{Sh}_p(G,X)$ *has canonical good reduction at every prime* $\mathfrak{p}$ *of* $E(G,X)$ *dividing* p, .

REMARK 16.5. Let $E_\mathfrak{p}$ be the completion of E at $\mathfrak{p}$, let $\hat{\mathcal{O}}_\mathfrak{p}$ be the ring of integers in $E_\mathfrak{p}$, and let $k(\mathfrak{p})$ be the residue field $\hat{\mathcal{O}}_\mathfrak{p}/\mathfrak{p}$.

(a) By $\mathrm{Sh}_p(G,X)$ having good reduction $\mathfrak{p}$, we mean that the inverse system

$$(\mathrm{Sh}_{K^pK_p}(G,X))_{K^p}, \quad K^p \subset G(\mathbb{A}_f^p) \text{ compact open, } K_p \text{ fixed,}$$

extends to an inverse system of flat schemes $\mathcal{S}_p = (\mathcal{S}_{K^p})$ over $\hat{\mathcal{O}}_\mathfrak{p}$ whose reduction modulo $\mathfrak{p}$ is an inverse system of varieties $(\overline{\mathrm{Sh}}_{K^pK_p}(G,X))_{K^p}$ over $k(\mathfrak{p})$ such that, for $K^p \supset K'^p$ sufficiently small,

$$\overline{\mathrm{Sh}}_{K^pK_p} \leftarrow \overline{\mathrm{Sh}}_{K'^pK_p}$$

is an étale map of smooth varieties. We require also that the action of $G(\mathbb{A}_f^p)$ on Sh_p extends to an action on $\mathcal{S}_p$.

(b) A variety over $E_\mathfrak{p}$ may not have good reduction to a smooth variety over $k(\mathfrak{p})$ — this can already be seen for elliptic curves — and, when it does it will generally have good reduction to many different smooth varieties, none of which is obviously the best. For example, given one good reduction, one can obtain another by blowing up a point in its closed fibre. By $\mathrm{Sh}_p(G,X)$ having *canonical* good reduction at $\mathfrak{p}$, I mean that, for any formally smooth scheme T over $\hat{\mathcal{O}}_\mathfrak{p}$,

$$\mathrm{Hom}_{\hat{\mathcal{O}}_\mathfrak{p}}(T, \varprojlim_{K^p} \mathcal{S}_{K^p}) \cong \mathrm{Hom}_{E_\mathfrak{p}}(T_{E_\mathfrak{p}}, \varprojlim_{K^p} \mathrm{Sh}_{K^pK_p}). \tag{61}$$

A smooth scheme is formally smooth, and an inverse limit of schemes étale over a smooth scheme is formally smooth. As $\varprojlim \mathcal{S}_{K^p}$ is formally smooth over $\hat{\mathcal{O}}_\mathfrak{p}$, (61) characterizes it uniquely up to a unique isomorphism (by the Yoneda lemma).

(c) In the Siegel case, Theorem 16.4 was proved by Mumford (his Fields medal theorem; Mumford 1965). In this case, the $\mathcal{S}_{K^p}$ and $\overline{\mathrm{Sh}}_{K^pK_p}$ are moduli schemes. The PEL-case can be considered folklore in that several authors have deduced it from the Siegel case and published sketches of proof, the most convincing of which is in Kottwitz 1992. In this case, $\mathcal{S}_p(G,X)$ is the zariski closure of $\mathrm{Sh}_p(G,X)$ in $\mathcal{S}_p(G(\psi),X(\psi))$ (cf. 5.16), and it is a moduli scheme. The hodge case[19] was proved by Vasiu (1999) except for a small set of primes. In this case, $\mathcal{S}_p(G,X)$ is the normalization of the zariski closure of $\mathrm{Sh}_p(G,X)$ in $\mathcal{S}_p(G(\psi),X(\psi))$. The case of abelian type follows easily from the hodge case.

(d) That Sh_p should have good reduction when K_p is hyperspecial was conjectured in Langlands 1976, p411. That there should be a canonical model characterized by a condition like that in (b) was conjectured in Milne 1992, §2.

[19]Over the reflex field, Shimura varieties of hodge type are no more difficult than Shimura varieties of PEL-type, but when one reduces modulo a prime they become much more difficult for two reasons: general tensors are more difficult to work with than endomorphisms, and little is known about hodge tensors in characteristic p.

Definition of the Langlands-Rapoport set. Let (G,X) be a Shimura datum for which SV4,5,6 hold, and let

$$\mathrm{Sh}_p(\mathbb{C}) = \mathrm{Sh}(\mathbb{C})/K_p = \varprojlim_{K^p} \mathrm{Sh}_{K^pK_p}(G,X)(\mathbb{C}).$$

For $x \in X$, let $I(x)$ be the subgroup $G(\mathbb{Q})$ fixing x, and let

$$S(x) = I(x)\backslash X^p(x) \times X_p(x), \quad X^p(x) = G(\mathbb{A}_f^p), \quad X_p(x) = G(\mathbb{Q}_p)/K_p.$$

One sees easily that there is a canonical bijection of sets with $G(\mathbb{A}_f^p)$-action

$$\bigsqcup S(x) \to \mathrm{Sh}_p(\mathbb{C})$$

where the left hand side is the disjoint union over a set of representatives for $G(\mathbb{Q})\backslash X$. This decomposition has a modular interpretation. For example, in the case of a Shimura variety of hodge type, the set $S(x)$ classifies the family of isomorphism classes of triples $(A,(s_i),\eta K)$ with $(A,(s_i))$ isomorphic to a fixed pair.

Langlands and Rapoport (1987, 5e) conjecture that $\overline{\mathrm{Sh}}_p(\mathbb{F})$ has a similar description except that now the left hand side runs over a set of isomorphism classes of homomorphisms $\phi\colon \mathfrak{P} \to E_G$. Recall that an isomorphism from one ϕ to a second ϕ' is an element g of $G(\mathbb{Q}^{\mathrm{al}})$ such that

$$\phi'(p) = g \cdot \phi(p) \cdot g^{-1}, \quad \text{all } p \in \mathfrak{P}.$$

Such a ϕ should be thought of as a "pre fake abelian motive with tensors". Specifically, if we fix a faithful representation $G \hookrightarrow \mathrm{GL}(V)$ and tensors t_i for V such that G is the subgroup of $\mathrm{GL}(V)$ fixing the t_i, then each ϕ gives a representation $\mathfrak{P} \to \mathrm{GL}(V(\mathbb{Q}^{\mathrm{al}})) \rtimes \Gamma$ (i.e., a fake abelian motive) plus tensors.

Definition of the set $S(\phi)$. We now fix a homomorphism $\phi\colon \mathfrak{P} \to E_G$ and define a set $S(\phi)$ equipped with a right action of $G(\mathbb{A}_f^p)$ and a commuting Frobenius operator Φ.

Definition of the group $I(\phi)$. The group $I(\phi)$ is defined to be the group of automorphisms of ϕ,

$$I(\phi) = \{g \in G(\mathbb{Q}^{\mathrm{al}}) \mid \mathrm{ad}(g) \circ \phi = \phi\}.$$

Note that if $\rho\colon G \to \mathrm{GL}(V)$ is a faithful representation of G, then $\rho \circ \phi$ is a fake motive and $I(\phi) \subset \mathrm{Aut}(\rho \circ \phi)$ (here we have abbreviated $\rho \rtimes 1$ to ρ).

Definition of $X^p(\phi)$. Let ℓ be a prime $\neq p, \infty$. We choose a prime w_ℓ of $\mathbb{Q}^{\mathrm{al}}$ dividing ℓ, and define $\mathbb{Q}_\ell^{\mathrm{al}}$ and $\Gamma_\ell \subset \Gamma$ as on p364.

Regard Γ_ℓ as an $\mathbb{Q}_\ell^{\mathrm{al}}/\mathbb{Q}_\ell$-affine extension with trivial kernel, and write ξ_ℓ for the homomorphism

$$\sigma \mapsto (1,\sigma)\colon \Gamma_\ell \to E_G(\ell), \quad E_G(\ell) = G(\mathbb{Q}_\ell^{\mathrm{al}}) \rtimes \Gamma_\ell.$$

From ϕ we get a homomorphism $\phi(\ell)\colon \mathfrak{P}(\ell) \to E_G(\ell)$, and, on composing this with the homomorphism $\zeta_\ell\colon \Gamma_\ell \to \mathfrak{P}(\ell)$ provided by (15.13), we get a second homomorphism $\Gamma_\ell \to E_G(\ell)$.

Define

$$X_\ell(\phi) = \mathrm{Isom}(\xi_\ell, \zeta_\ell \circ \phi(\ell)).$$

Clearly, $\mathrm{Aut}(\xi_\ell) = G(\mathbb{Q}_\ell)$ acts on $X_\ell(\phi)$ on the right, and $I(\phi)$ acts on the left. If $X_\ell(\phi)$ is nonempty, then the first action makes $X_\ell(\phi)$ into a principal homogeneous space for $G(\mathbb{Q}_\ell)$.

Note that if $\rho\colon G \to \mathrm{GL}(V)$ is a faithful representation of G, then

$$X_\ell(\phi) \subset \mathrm{Isom}(V(\mathbb{Q}_\ell), V_\ell(\rho \circ \phi)). \tag{62}$$

By choosing the ζ_ℓ judiciously (cf. p365), we obtain compact open subspaces of the $X_\ell(\phi)$, and we can define $X^p(\phi)$ to be the restricted product of the $X_\ell(\phi)$. If nonempty, it is a principal homogeneous space for $G(\mathbb{A}_f^p)$.

Definition of $X_p(\phi)$. We choose a prime w_p of $\mathbb{Q}^{\mathrm{al}}$ dividing p, and we use the notations of p365. We let L denote the completion of $\mathbb{Q}_p^{\mathrm{un}}$, and we let $\mathcal{O}_L$ denote the ring of integers in L (it is the ring of Witt vectors with coefficients in $\mathbb{F}$). We let σ Frobenius automorphism of $\mathbb{Q}_p^{\mathrm{un}}$ or L that acts as $x \mapsto x^p$ on the residue field.

From ϕ and (15.14), we have homomorphisms

$$D \longrightarrow \mathfrak{P}(p)^{\mathrm{un}} \xrightarrow{\phi(p)^{\mathrm{un}}} G(\mathbb{Q}_p^{\mathrm{un}}) \rtimes \Gamma_p^{\mathrm{un}}.$$

For some n, the composite factors through D_n. There is a canonical element in D_n mapping to σ, and we let (b,σ) denote its image in $G(\mathbb{Q}_p^{\mathrm{un}}) \rtimes \Gamma_p^{\mathrm{un}}$. The image $b(\phi)$ of b in $G(L)$ is well-defined up to σ-conjugacy, i.e., if $b(\phi)'$ also arises in this way, then $b(\phi)' = g^{-1} \cdot b(\phi) \cdot \sigma g$.

Note that if $\rho\colon G \to \mathrm{GL}(V)$ is a faithful representation of G, then $D(\phi \circ \rho)$ is $V(L)$ with F acting as $v \mapsto b(\phi)\sigma v$.

Recall p344 that we have a well-defined $G(\mathbb{Q}^{\mathrm{al}})$-conjugacy class $c(X)$ of cocharacters of $G_{\mathbb{Q}^{\mathrm{al}}}$. We can transfer this to conjugacy class of cocharacters of $G_{\mathbb{Q}_p^{\mathrm{al}}}$, which contains an element μ defined over $\mathbb{Q}_p^{\mathrm{un}}$ (see 12.3; G splits over $\mathbb{Q}_p^{\mathrm{un}}$ because we are assuming it contains a hyperspecial group). Let

$$C_p = G(\mathcal{O}_L) \cdot \mu(p) \cdot G(\mathcal{O}_L).$$

Here we are writing $G(\mathcal{O}_L)$ for $\mathcal{G}(\mathcal{O}_L)$ with $\mathcal{G}$ as in the definition of hyperspecial.

Define

$$X_p(\phi) = \{g \in G(L)/G(\mathcal{O}_L) \mid g^{-1} \cdot b(\phi) \cdot g \in C_p\}.$$

There is a natural action of $I(\phi)$ on this set.

Definition of the Frobenius element Φ. For $g \in X_p(\phi)$, define

$$\Phi(g) = b(\phi) \cdot \sigma b(\phi) \cdot \dots \cdot \sigma^{m-1} b(\phi) \cdot \sigma^m g$$

where $m = [E_v\colon \mathbb{Q}_p]$.

The set $S(\phi)$. Let

$$S(\phi) = I(\phi)\backslash X^p(\phi) \times X_p(\phi).$$

Since $I(\phi)$ acts on both $X^p(\phi)$ and $X_p(\phi)$, this makes sense. The group $G(\mathbb{A}_f^p)$ acts on $S(\phi)$ through its action on $X^p(\phi)$ and Φ acts through its action on $X_p(\phi)$.

The admissibility condition. The homomorphisms $\phi\colon \mathfrak{P} \to E_G$ contributing to the Langlands-Rapoport set must satisfy an admissibility condition at each prime plus one global condition.

The condition at ∞. Let E_∞ be the extension

$$1 \to \mathbb{C}^\times \to E_\infty \to \Gamma_\infty \to 1, \quad \Gamma_\infty = \mathrm{Gal}(\mathbb{C}/\mathbb{R}) = \langle \iota \rangle$$

associated with the quaternion algebra $\mathbb{H}$, and regard it as an affine extension with kernel $\mathbb{G}_m$. Note that $E_\infty = \mathbb{C}^\times \sqcup \mathbb{C}^\times j$ and $jzj^{-1} = \bar{z}$.

From the diagram (60) with $l = \infty$, we obtain a $\mathbb{C}/\mathbb{R}$-affine extension

$$1 \to P(\mathbb{C}) \to \mathfrak{P}(\infty) \to \Gamma_\infty \to 1.$$

LEMMA 16.6. *There is a homomorphism $\zeta_\infty\colon E_\infty \to \mathfrak{P}(\infty)$ whose restriction to the kernels, $\mathbb{G}_m \mapsto P_{\mathbb{C}}$, corresponds to the map on characters $\pi \mapsto wt(\pi)$.*

PROOF. This follows from the fact that the homomorphism $H^2(\Gamma_\infty, \mathbb{G}_m) \to H^2(\Gamma_\infty, P_\mathbb{R})$ sends the cohomology class of E_∞ to that of $\mathfrak{P}(\infty)$. □

LEMMA 16.7. *For any $x \in X$, the formulas*

$$\xi_x(z) = (w_X(z), 1), \quad \xi_x(j) = (\mu_x(-1)^{-1}, \iota)$$

define a homomorphism $E_\infty \to \mathfrak{P}(\infty)$. Replacing x with a different point, replaces the homomorphism with an isomorphic homomorphism.

PROOF. Easy exercise. □

Write ξ_X for the isomorphism class of homomorphisms defined in (16.7). Then the admissibility condition at ∞ is that $\zeta_\infty \circ \phi(\infty) \in \xi_X$.

The condition at $\ell \neq p$. The admissibility condition at $\ell \neq p$ is that the set $X_\ell(\phi)$ be nonempty, i.e., that $\zeta_\ell \circ \phi(\ell)$ be isomorphic to ξ_ℓ.

The condition at p. The condition at p is that the set $X_p(\phi)$ be nonempty.

The global condition. Let $\nu\colon G \to T$ be the quotient of G by its derived group. From X we get a conjugacy class of cocharacters of $G_\mathbb{C}$, and hence a well defined cocharacter μ of T. Under our hypotheses on (G, X), μ satisfies the conditions of (15.15), and so defines a homomorphism $\phi_\mu\colon \mathfrak{P} \to E_T$. The global condition is that $\nu \circ \phi$ be isomorphic to ϕ_μ.

The Langlands-Rapoport set. The Langlands-Rapoport set

$$\mathrm{LR}(G, X) = \bigsqcup S(\phi)$$

where the disjoint union is over a set of representatives for the isomorphism classes of admissible homomorphism $\phi\colon \mathfrak{P} \to E_G$. There are commuting actions of $G(\mathbb{A}_f^p)$ and of the Frobenius operator Φ on $\mathrm{LR}(G, X)$.

The conjecture of Langlands and Rapoport.

CONJECTURE 16.8 (Langlands and Rapoport 1987). *Let (G, X) be a Shimura datum satisfying SV4,5,6 and such that G^{der} is simply connected, and let K_p be a hyperspecial subgroup of $G(\mathbb{Q}_p)$. Let $\mathfrak{p}$ be a prime of $E(G, X)$ dividing p, and assume that* Sh_p *has canonical good reduction at $\mathfrak{p}$. Then there is a bijection of sets*

$$\mathrm{LR}(G, X) \to \overline{\mathrm{Sh}}_p(G, X)(\mathbb{F}) \tag{63}$$

compatible with the actions $G(\mathbb{A}_f^p)$ and the Frobenius elements.

REMARK 16.9. (a) The conditions SV5 and SV6 are not in the original conjecture — I included them to simplify the statement of the conjecture.

(b) There is also a conjecture in which one does not require SV4, but this requires that $\mathfrak{P}$ be replaced by a more complicated affine extension $\mathfrak{Q}$.

(c) The conjecture as originally stated is definitely wrong without the assumption that G^{der} is simply connected. However, when one replaces the "admissible homomorphisms" in the statement with another notion, that of "special homomorphisms", one obtains a statement that should be true for all Shimura varieties. In fact, it is known that the statement with G^{der} simply connected then implies the general statement (see Milne 1992, §4, for the details and a more precise statement).

(d) It is possible to state, and prove, a conjecture similar to (16.8) for zero-dimensional Shimura varieties. The map $(G, X) \to (T, Y)$ (see p311) defines a map of the associated Langlands-Rapoport sets, and we should add to the conjecture that

$$\begin{tikzcd} LR(G,X) \arrow[r] \arrow[d] & \mathrm{Sh}_p(G,X)(\mathbb{F}) \arrow[d] \\ LR(T,Y) \arrow[r] & \mathrm{Sh}_p(T,Y)(\mathbb{F}) \end{tikzcd}$$

commutes.

17. A formula for the number of points

A reader of the last two sections may be sceptical of the value of a description like (63), even if proved. In this section we briefly explain how it leads to a very explicit, and useful, formula for the number of points on the reduction of a Shimura variety with values in a finite field.

Throughout, (G,X) is a Shimura datum satisfying SV4,5,6 and K_p is a hyperspecial subgroup of $G(\mathbb{Q}_p)$. We assume that G^{der} simply connected and that $\mathrm{Sh}_p(G,X)$ has canonical good reduction at a prime $\mathfrak{p}|p$ of the reflex field $E = E(G,X)$. Other notations are as in the last section; for example, L_n is the subfield of $\mathbb{Q}_p^{\mathrm{un}}$ of degree n over $\mathbb{Q}_p$ and L is the completion of $\mathbb{Q}_p^{\mathrm{un}}$. We fix a field $\mathbb{F}_q \supset k(\mathfrak{p}) \supset \mathbb{F}_p$, $q = p^n$.

Triples. We consider triples $(\gamma_0;\gamma,\delta)$ where

- γ_0 is a semisimple element of $G(\mathbb{Q})$ that is contained in an elliptic torus of $G_{\mathbb{R}}$ (i.e., a torus that is anisotropic modulo the centre of $G_{\mathbb{R}}$),
- $\gamma = (\gamma(\ell))_{\ell\neq p,\infty}$ is an element of $G(\mathbb{A}_f^p)$ such that, for all ℓ, $\gamma(\ell)$ becomes conjugate to γ_0 in $G(\mathbb{Q}_\ell^{\mathrm{al}})$,
- δ is an element of $G(L_n)$ such that
$$\mathcal{N}\delta \stackrel{\mathrm{df}}{=} \delta\cdot\sigma\delta\cdot\ldots\cdot\sigma^{n-1}\delta,$$
becomes conjugate to γ_0 in $G(\mathbb{Q}_p^{\mathrm{al}})$.

Two triples $(\gamma_0;\gamma,\delta)$ and $(\gamma_0';\gamma',\delta')$ are said to be ***equivalent***, $(\gamma_0;\gamma,\delta)\sim(\gamma_0';\gamma',\delta')$, if γ_0 is conjugate to γ_0' in $G(\mathbb{Q})$, $\gamma(\ell)$ is conjugate to $\gamma'(\ell)$ in $G(\mathbb{Q}_\ell)$ for each $\ell\neq p,\infty$, and δ is σ-conjugate to δ' in $G(L_n)$.

Given such a triple $(\gamma_0;\gamma,\delta)$, we set:

- $I_0 = G_{\gamma_0}$, the centralizer of γ_0 in G; it is connected and reductive;
- $I_\infty =$ the inner form of $I_{0\mathbb{R}}$ such that $I_\infty/Z(G)$ is anisotropic;
- $I_\ell =$ the centralizer of $\gamma(\ell)$ in $G_{\mathbb{Q}_\ell}$;
- $I_p =$ the inner form of $G_{\mathbb{Q}_p}$ such that $I_p(\mathbb{Q}_p) = \{x\in G(L_n)\mid x^{-1}\cdot\delta\cdot\sigma x = \delta\}$.

We need to assume that the triple satisfies the following condition:

(*) there exists an inner form I of I_0 such that $I_{\mathbb{Q}_\ell}$ is isomorphic to I_ℓ for all ℓ (including p and ∞).

Because γ_0 and γ_ℓ are stably conjugate, there exists an isomorphism $a_\ell\colon I_{0,\mathbb{Q}_\ell^{\mathrm{al}}}\to I_{\ell,\mathbb{Q}_\ell^{\mathrm{al}}}$, well-defined up to an inner automorphism of I_0 over $\mathbb{Q}_\ell^{\mathrm{al}}$. Choose a system $(I,a,(j_\ell))$ consisting of a $\mathbb{Q}$-group I, an inner twisting $a\colon I_0\to I$ (isomorphism over $\mathbb{Q}^{\mathrm{al}}$), and isomorphisms $j_\ell\colon I_{\mathbb{Q}_\ell}\to I_\ell$ over $\mathbb{Q}_\ell$ for all ℓ, unramified for almost all ℓ, such that $j_\ell\circ a$ and a_ℓ differ by an inner automorphism — our assumption

(*) guarantees the existence of such a system. Moreover, any other such system is isomorphic to one of the form $(I, a, (j_\ell \circ \mathrm{ad} h_\ell))$ where $(h_\ell) \in I^{\mathrm{ad}}(\mathbb{A})$.

Let dx denote the Haar measure on $G(\mathbb{A}_f^p)$ giving measure 1 to K^p. Choose a Haar measure di^p on $I(\mathbb{A}_f^p)$ that gives rational measure to compact open subgroups of $I(\mathbb{A}_f^p)$, and use the isomorphisms j_ℓ to transport it to a measure on $G(\mathbb{A}_f^p)_\gamma$ (the centralizer of γ in $G(\mathbb{A}_f^p)$). The resulting measure does not change if (j_ℓ) is modified by an element of $I^{\mathrm{ad}}(\mathbb{A})$. Write $d\overline{x}$ for the quotient of dx by di^p. Let f be an element of the Hecke algebra $\mathcal{H}$ of locally constant K-bi-invariant $\mathbb{Q}$-valued functions on $G(\mathbb{A}_f)$, and assume that $f = f^p \cdot f_p$ where f^p is a function on $G(\mathbb{A}_f^p)$ and f_p is the characteristic function of K_p in $G(\mathbb{Q}_p)$ divided by the measure of K_p. Define

$$O_\gamma(f^p) = \int_{G(\mathbb{A}_f^p)_\gamma \backslash G(\mathbb{A}_f^p)} f^p(x^{-1}\gamma x)\, d\overline{x}$$

Let dy denote the Haar measure on $G(L_n)$ giving measure 1 to $G(\mathcal{O}_{L_n})$. Choose a Haar measure di_p on $I(\mathbb{Q}_p)$ that gives rational measure to the compact open subgroups, and use j_p to transport the measure to $I_p(\mathbb{Q}_p)$. Again the resulting measure does not change if j_p is modified by an element of $I^{\mathrm{ad}}(\mathbb{Q}_p)$. Write $d\overline{y}$ for the quotient of dy by di_p. Proceeding as on p370, we choose a cocharacter μ in $c(X)$ well-adapted to the hyperspecial subgroup K_p and defined over L_n, and we let φ be the characteristic function of the coset $G(\mathcal{O}_{L_n}) \cdot \mu(p) \cdot G(\mathcal{O}_{L_n})$. Define

$$TO_\delta(\varphi) = \int_{I(\mathbb{Q}_p)\backslash G(L_n)} \varphi(y^{-1}\delta\sigma(y))d\overline{y}$$

Since $I/Z(G)$ is anisotropic over $\mathbb{R}$, and since we are assuming SV5, $I(\mathbb{Q})$ is a discrete subgroup of $I(\mathbb{A}_f^p)$, and we can define the volume of $I(\mathbb{Q})\backslash I(\mathbb{A}_f)$. It is a rational number because of our assumption on di^p and di_p. Finally, define

$$I(\gamma_0; \gamma, \delta) = I(\gamma_0; \gamma, \delta)(f^p, r) = \mathrm{vol}(I(\mathbb{Q})\backslash I(\mathbb{A}_f)) \cdot O_\gamma(f^p) \cdot TO_\delta(\phi_r).$$

The integral $I(\gamma_0; \gamma, \delta)$ is independent of the choices made, and

$$(\gamma_0; \gamma, \delta) \sim (\gamma_0'; \gamma', \delta') \implies I(\gamma_0; \gamma, \delta) = I(\gamma_0'; \gamma', \delta').$$

The triple attached to an admissible pair (ϕ, ε). An ***admissible pair*** (ϕ, γ_0) is an admissible homomorphism $\phi\colon \mathfrak{P} \to E_G$ and a $\gamma \in I_\phi(\mathbb{Q})$ such that $\gamma_0 x = \Phi^r x$ for some $x \in X_p(\phi)$. Here $r = [k(\mathfrak{p})\colon \mathbb{F}_p]$. An ***isomorphism*** $(\phi, \gamma_0) \to (\phi', \gamma_0')$ of admissible pairs is an isomorphism $\phi \to \phi'$ sending γ to γ', i.e., it is a $g \in G(\mathbb{Q}^{\mathrm{al}})$ such that

$$\mathrm{ad}(g) \circ \phi = \phi', \quad \mathrm{ad}(g)(\gamma) = \gamma'.$$

Let $(T, x) \subset (G, X)$ be a special pair. Because of our assumptions on (G, X), the cocharacter μ_x of T satisfies the conditions of (15.15) and so defines a homomorphism $\phi_x\colon \mathfrak{P} \to E_T$. Langlands and Rapoport (1987, 5.23) show that every admissible pair is isomorphic to a pair (ϕ, γ) with $\phi = \phi_x$ and $\gamma \in T(\mathbb{Q})$. For such a pair (ϕ, γ), $b(\phi)$ is represented by a $\delta \in G(L_n)$ which is well-defined up to conjugacy.

Let γ be the image of γ_0 in $G(\mathbb{A}_f^p)$. Then the triple $(\gamma_0; \gamma, \delta)$ satisfies the conditions in the last subsection. A triple arising in this way from an admissible pair will be called ***effective.***

The formula. For a triple $(\gamma_0 \ldots)$, the kernel of

$$H^1(\mathbb{Q}, I_0) \to H^1(\mathbb{Q}, G) \oplus \prod_l H^1(\mathbb{Q}_l, I_0)$$

is finite — we denote its order by $c(\gamma_0)$.

THEOREM 17.1. *Let (G, X) be a Shimura datum satisfying the hypotheses of (16.8). If that conjecture is true, then*

$$\# \operatorname{Sh}_p(\mathbb{F}_q) = \sum_{(\gamma_0;\gamma,\delta)} c(\gamma_0) \cdot I(\gamma_0; \gamma, \delta) \tag{64}$$

where the sum is over a set of representatives for the effective triples.

PROOF. See Milne 1992, 6.13. □

NOTES. Early versions of (64) can be found in papers of Langlands, but the first precise general statement of such a formula is in Kottwitz 1990. There Kottwitz attaches a cohomological invariant $\alpha(\gamma_0; \gamma, \delta)$ to a triple $(\gamma_0; \gamma, \delta)$, and conjectures that the formula (64) holds if the sum is taken over a set of representatives for the triples with $\alpha = 1$ (ibid. §3). Milne (1992, 7.9) proves that, among triples contributing to the sum, $\alpha = 1$ if and only if the triple is effective, and so the conjecture of Langlands and Rapoport implies Kottwitz's conjecture.[20] On the other hand, Kottwitz (1992) proves his conjecture for Shimura varieties of simple PEL type A or C unconditionally (without however proving the conjecture of Langlands and Rapoport for these varieties).

References

Ash, A., Mumford, D., Rapoport, M. and Tai, Y. (1975), *Smooth compactification of locally symmetric varieties*, Math. Sci. Press, Brookline, Mass.

Baily, Jr., W. L. and Borel, A. (1966), 'Compactification of arithmetic quotients of bounded symmetric domains', *Ann. of Math. (2)* **84**, 442–528.

Baily, W. L. (1958), 'Satake's compactification of V_n', *Amer. J. Math.* **80**, 348–364.

Boothby, W. M. (1975), *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press, New York-London.

Borel, A. (1962), Ensembles fondamentaux pour les groupes arithmétiques, *in* 'Colloq. Théorie des Groupes Algébriques (Bruxelles, 1962)', Librairie Universitaire, Louvain, pp. 23–40.

Borel, A. (1969), *Introduction aux groupes arithmétiques*, Publications de l'Institut de Mathématique de l'Université de Strasbourg, XV. Actualités Scientifiques et Industrielles, No. 1341, Hermann, Paris.

Borel, A. (1972), 'Some metric properties of arithmetic quotients of symmetric spaces and an extension theorem', *J. Differential Geometry* **6**, 543–560.

Borel, A. (1991), *Linear algebraic groups*, Vol. 126 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Borel, A. (1998), *Semisimple groups and Riemannian symmetric spaces*, Vol. 16 of *Texts and Readings in Mathematics*, Hindustan Book Agency, New Delhi.

Borel, A. and Harish-Chandra (1962), 'Arithmetic subgroups of algebraic groups', *Ann. of Math. (2)* **75**, 485–535.

Borovoĭ, M. V. (1984), 'Langlands' conjecture concerning conjugation of connected Shimura varieties', *Selecta Math. Soviet.* **3**(1), 3–39.

Borovoĭ, M. V. (1987), The group of points of a semisimple group over a totally real-closed field, *in* 'Problems in group theory and homological algebra (Russian)', Matematika, Yaroslav. Gos. Univ., pp. 142–149.

[20]At least in the case that the weight is rational — Kottwitz does not make this assumption.

Bourbaki, N. (1958), *Éléments de mathématique. I: Les structures fondamentales de l'analyse. Fascicule VII. Livre II: Algèbre. Chapitre 3: Algèbre multilinéaire*, Nouvelle édition. Actualités Scientifiques et Industrielles, No. 1044, Hermann, Paris.

Bourbaki, N. (1981), *Groupes et Algèbres de Lie. Chapitres 4,5 et 6*, Elements of Mathematics, Masson, Paris.

Bourbaki, N. (1989), *General topology. Chapters 1–4*, Elements of Mathematics, Springer-Verlag, Berlin.

Bravo, A., Encinas, S. and Villamayor, O. (2002), 'A Simplified Proof of Desingularization and Applications'. arXiv:math.AG/0206244.

Brylinski, J.-L. (1983), "1-motifs" et formes automorphes (théorie arithmétique des domaines de Siegel), *in* 'Conference on automorphic theory (Dijon, 1981)', Vol. 15 of *Publ. Math. Univ. Paris VII*, Univ. Paris VII, Paris, pp. 43–106.

Cartan, H. (1958), 'Prolongement des espaces analytiques normaux', *Math. Ann.* **136**, 97–110.

Cartan, H. (1963), *Elementary theory of analytic functions of one or several complex variables*, Éditions Scientifiques Hermann, Paris.

Chow, W.-L. (1949), 'On compact complex analytic varieties', *Amer. J. Math.* **71**, 893–914.

Deligne, P. (1970), *Équations différentielles à points singuliers réguliers*, Springer-Verlag, Berlin.

Deligne, P. (1971*a*), 'Théorie de Hodge. II', *Inst. Hautes Études Sci. Publ. Math.* **40**, 5–57.

Deligne, P. (1971*b*), Travaux de Shimura, *in* 'Séminaire Bourbaki, 23ème année (1970/71), Exp. No. 389', Springer, Berlin, pp. 123–165. Lecture Notes in Math., Vol. 244.

Deligne, P. (1971*c*), 'Travaux de Shimura'. Preliminary unpublished version of Deligne 1971b.

Deligne, P. (1973*a*), *Espaces hermitiens symétriques, Spring 1973*, Transcription available at www.jmilne.org/math/.

Deligne, P. (1973*b*), Les constantes des équations fonctionnelles des fonctions L, *in* 'Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972)', Springer, Berlin, pp. 501–597. Lecture Notes in Math., Vol. 349.

Deligne, P. (1979), Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques, *in* 'Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2', Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., pp. 247–289.

Deligne, P. (1982), Hodge cycles on abelian varieties (notes by J.S. Milne), *in* 'Hodge cycles, motives, and Shimura varieties', Vol. 900 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, pp. 9–100. Available at www.jmilne.org/math/.

Demazure, M. (1972), *Lectures on p-divisible groups*, Springer-Verlag, Berlin.

Farb, B. and Dennis, R. K. (1993), *Noncommutative algebra*, Vol. 144 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Giraud, J. (1968), 'Remarque sur une formule de Shimura-Taniyama', *Invent. Math.* **5**, 231–236.

Goresky, M. (2003), 'Compactifications of modular varieties'. Preliminary version of article for this volume.

Greenberg, M. J. (1967), *Lectures on algebraic topology*, W. A. Benjamin, Inc., New York-Amsterdam.

Hartshorne, R. (1977), *Algebraic geometry*, Springer-Verlag, New York.

Hatcher, A. (2002), *Algebraic topology*, Cambridge University Press, Cambridge. Available at http://www.math.cornell.edu/ hatcher/.

Helgason, S. (1978), *Differential geometry, Lie groups, and symmetric spaces*, Vol. 80 of *Pure and Applied Mathematics*, Academic Press Inc., New York.

Hironaka, H. (1964), 'Resolution of singularities of an algebraic variety over a field of characteristic zero. I, II', *Ann. of Math. (2) 79 (1964), 109–203; ibid. (2)* **79**, 205–326.

Humphreys, J. E. (1972), *Introduction to Lie algebras and representation theory*, Springer-Verlag, New York.

Humphreys, J. E. (1978), *Introduction to Lie algebras and representation theory*, Vol. 9 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Ihara, Y. (1968), *On congruence monodromy problems. Vol. 1*, Lecture Notes, No. 1, Department of Mathematics, University of Tokyo, Tokyo.

Iitaka, S. (1982), *Algebraic geometry*, Vol. 76 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Knus, M.-A., Merkurjev, A., Rost, M. and Tignol, J.-P. (1998), *The book of involutions*, Vol. 44 of *American Mathematical Society Colloquium Publications*, American Mathematical Society, Providence, RI.

Kottwitz, R. E. (1984), 'Shimura varieties and twisted orbital integrals', *Math. Ann.* **269**(3), 287–300.

Kottwitz, R. E. (1990), Shimura varieties and λ-adic representations, *in* 'Automorphic forms, Shimura varieties, and L-functions, Vol. I (Ann Arbor, MI, 1988)', Vol. 10 of *Perspect. Math.*, Academic Press, Boston, MA, pp. 161–209.

Kottwitz, R. E. (1992), 'Points on some Shimura varieties over finite fields', *J. Amer. Math. Soc.* **5**(2), 373–444.

Krantz, S. G. (1982), *Function theory of several complex variables*, John Wiley & Sons Inc., New York.

Kwack, M. H. (1969), 'Generalization of the big Picard theorem', *Ann. of Math. (2)* **90**, 9–22.

Langlands, R. P. (1976), Some contemporary problems with origins in the Jugendtraum, *in* 'Mathematical developments arising from Hilbert problems (Proc. Sympos. Pure Math., Vol. XXVIII, Northern Illinois Univ., De Kalb, Ill., 1974)', Amer. Math. Soc., Providence, R. I., pp. 401–418.

Langlands, R. P. (1977), 'Shimura varieties and the Selberg trace formula', *Canad. J. Math.* **29**(6), 1292–1299.

Langlands, R. P. and Rapoport, M. (1987), 'Shimuravarietäten und Gerben', *J. Reine Angew. Math.* **378**, 113–220.

Lee, J. M. (1997), *Riemannian manifolds*, Vol. 176 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Mac Lane, S. (1998), *Categories for the working mathematician*, Vol. 5 of *Graduate Texts in Mathematics*, second edn, Springer-Verlag, New York.

Matsumoto, H. (1969), 'Sur les sous-groupes arithmétiques des groupes semi-simples déployés', *Ann. Sci. École Norm. Sup. (4)* **2**, 1–62.

Milne, J. S. (1979), Complex multiplication for pedestrians. September 19, 1979; revised June 7, 1998, available at www.jmilne.org/math/.

Milne, J. S. (1983), The action of an automorphism of **C** on a Shimura variety and its special points, *in* 'Arithmetic and geometry, Vol. I', Vol. 35 of *Progr. Math.*, Birkhäuser Boston, Boston, MA, pp. 239–265.

Milne, J. S. (1986), Abelian varieties, *in* 'Arithmetic geometry (Storrs, Conn., 1984)', Springer, New York, pp. 103–150.

Milne, J. S. (1987), The (failure of the) Hasse principle for centres of semisimple groups. 6th June, 1987, available at www.jmilne.org/math/.

Milne, J. S. (1992), The points on a Shimura variety modulo a prime of good reduction, *in* 'The zeta functions of Picard modular surfaces', Univ. Montréal, Montreal, QC, pp. 151–253.

Milne, J. S. (1994), Shimura varieties and motives, *in* 'Motives (Seattle, WA, 1991)', Vol. 55 of *Proc. Sympos. Pure Math.*, Amer. Math. Soc., Providence, RI, pp. 447–523.

Milne, J. S. (1999), 'Descent for Shimura varieties', *Michigan Math. J.* **46**(1), 203–208.

Milne, J. S. (2003), 'Gerbes and abelian motives'. Preprint available at www.jmilne.org/math/.

Milne, J. S. and Shih, K.-y. (1981), 'Automorphism groups of Shimura varieties and reciprocity laws', *Amer. J. Math.* **103**(5), 911–935.

Mok, N. and Zhong, J. Q. (1989), 'Compactifying complete Kähler-Einstein manifolds of finite topological type and bounded curvature', *Ann. of Math. (2)* **129**(3), 427–470.

Mumford, D. (1965), *Geometric invariant theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Band 34, Springer-Verlag, Berlin.

Mumford, D. (1969), 'A note of Shimura's paper "Discontinuous groups and abelian varieties"', *Math. Ann.* **181**, 345–351.

Mumford, D. (1970), *Abelian varieties*, Tata Institute of Fundamental Research Studies in Mathematics, No. 5, Published for the Tata Institute of Fundamental Research, Bombay.

Mumford, D. (1977), 'Hirzebruch's proportionality theorem in the noncompact case', *Invent. Math.* **42**, 239–272.

Murty, V. K. (1993), *Introduction to abelian varieties*, Vol. 3 of *CRM Monograph Series*, American Mathematical Society, Providence, RI.

Nadel, A. and Tsuji, H. (1988), 'Compactification of complete Kähler manifolds of negative Ricci curvature', *J. Differential Geom.* **28**(3), 503–512.

Platonov, V. and Rapinchuk, A. (1994), *Algebraic groups and number theory*, Vol. 139 of *Pure and Applied Mathematics*, Academic Press Inc., Boston, MA.

Raghunathan, M. S. (1972), *Discrete subgroups of Lie groups*, Springer-Verlag, New York.

Satake, I. (1956), 'On the compactification of the Siegel space', *J. Indian Math. Soc. (N.S.)* **20**, 259–281.

Satake, I. (1960), 'On compactifications of the quotient spaces for arithmetically defined discontinuous groups', *Ann. of Math. (2)* **72**, 555–580.

Satake, I. (1980), *Algebraic structures of symmetric domains*, Vol. 4 of *Kanô Memorial Lectures*, Iwanami Shoten, Tokyo.

Satake, I. (2001), 'Compactifications, old and new [translation of Sūgaku **51** (1999), no. 2, 129–141; MR 2000j:32035]', *Sugaku Expositions* **14**(2), 175–189.

Satake, I. and Cartan, H. (1958), Démonstration du théorème fondamental, *in* 'Séminaire Henri Cartan; 10e année: 1957/58. Fonctions Automorphes, Exposé 15', Secrétariat mathématique, ENS, Paris, p. 12.

Serre, J.-P. (1964), 'Sur les groupes de congruence des variétés abéliennes', *Izv. Akad. Nauk SSSR Ser. Mat.* **28**, 3–20.

Serre, J.-P. (1968), *Abelian l-adic representations and elliptic curves*, McGill University lecture notes written with the collaboration of Willem Kuyk and John Labute, W. A. Benjamin, Inc., New York-Amsterdam.

Serre, J.-P. and Tate, J. (1968), 'Good reduction of abelian varieties', *Ann. of Math. (2)* **88**, 492–517.

Shafarevich, I. R. (1994), *Basic algebraic geometry. 1,2*, Springer-Verlag, Berlin.

Shimura, G. (1963), 'On analytic families of polarized abelian varieties and automorphic functions', *Ann. of Math. (2)* **78**, 149–192.

Shimura, G. (1971), On arithmetic automorphic functions, *in* 'Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 2', Gauthier-Villars, Paris, pp. 343–348.

Shimura, G. and Taniyama, Y. (1961), *Complex multiplication of abelian varieties and its applications to number theory*, Vol. 6 of *Publications of the Mathematical Society of Japan*, The Mathematical Society of Japan, Tokyo.

Siegel, C. L. (1939), 'Einführung in die Theorie der Modulfunktionen n-ten Grades', *Math. Ann.* **116**, 617–657.

Springer, T. A. (1998), *Linear algebraic groups*, Vol. 9 of *Progress in Mathematics*, Birkhäuser Boston Inc., Boston, MA.

Tate, J. T. (1967), Global class field theory, *in* 'Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965)', Thompson, Washington, D.C., pp. 162–203.

Tate, J. T. (1969), 'Classes d'isogénie des variétés abéliennes sur un corps fini'. Séminaire Bourbaki.

Taylor, J. L. (2002), *Several complex variables with connections to algebraic geometry and Lie groups*, Vol. 46 of *Graduate Studies in Mathematics*, American Mathematical Society, Providence, RI.

Vasiu, A. (1999), 'Integral canonical models of Shimura varieties of preabelian type', *Asian J. Math.* **3**(2), 401–518.

Voisin, C. (2002), *Théorie de Hodge et géométrie algébrique complexe*, Vol. 10 of *Cours Spécialisés*, Société Mathématique de France, Paris. English translation available from Cambridge University Press.

Waterhouse, W. C. (1979), *Introduction to affine group schemes*, Vol. 66 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Weil, A. (1948), *Variétés abéliennes et courbes algébriques*, Actualités Sci. Ind., no. 1064 = Publ. Inst. Math. Univ. Strasbourg 8 (1946), Hermann & Cie., Paris.

Weil, A. (1956), 'The field of definition of a variety', *Amer. J. Math.* **78**, 509–524.

Weil, A. (1958), *Introduction à l'étude des variétés kählériennes*, Publications de l'Institut de Mathématique de l'Université de Nancago, VI. Actualités Sci. Ind. no. 1267, Hermann, Paris.

Wells, R. O. (1980), *Differential analysis on complex manifolds*, Vol. 65 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Whitney, H. (1957), 'Elementary structure of real algebraic varieties', *Ann. of Math. (2)* **66**, 545–556.

Witte, D. (2001), *Introduction to Arithmetic Groups*, arXiv:math.DG/0106063.

Wolf, J. A. (1984), *Spaces of constant curvature*, Publish or Perish Inc., Houston, TX.

Zink, T. (1983), 'Isogenieklassen von Punkten von Shimuramannigfaltigkeiten mit Werten in einem endlichen Körper', *Math. Nachr.* **112**, 103–124.

E-mail address: `math@jmilne.org`