Introduction to Shimura Varieties

J.S. Milne

October 23, 2004; revised September 16, 2017

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. In this revised version, the numbering is unchanged from the original published version except for displays.

Contents

These are my notes for a series of fifteen lectures at the Clay Summer School, Fields Institute, Toronto, June 2 – June 27, 2003. The original version was published as: "Introduction to Shimura varieties, In *Harmonic analysis, the trace formula, and Shimura varieties,* 265–378, Clay Math. Proc., 4, Amer. Math. Soc., Providence, RI, 2005." The notes were revised in 2017.

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 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 and his students 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 his 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.

The first nine sections study Shimura varieties over the complex numbers, the next five study them over number fields of characteristic zero (the theory of canonical models), and the final three study them in mixed characteristic and over finite fields.

Introduction to the revised version (2017)

On looking at these notes thirteen years after they were written, I found that they read too closely as being my personal notes for the lectures. In particular, they lacked the motivation and historical background that (I hope) the lectures provided. In revising them, I have added this background, and I have fixed all the errors and instances of careless writing that have been pointed out to me. Unnumbered asides are new, and this version includes three appendices not in the published version.

One point I should emphasize is that this is an introduction to the theory of *general* Shimura varieties. Although Shimura varieties of PEL-type form a very important class — they are the moduli varieties of abelian varieties with polarization, endomorphism, and level structure — they make up only a small class in the totality of Shimura varieties.[2]

The simplest Shimura varieties are the elliptic modular curves. My notes *Modular Functions and Modular Forms* emphasize the arithmetic and the geometry of these curves, and so provide an elementary preview of some of the theory discussed in these notes.

The entire foundations of the theory of Shimura varieties need to be reworked. Once that has been accomplished, perhaps I will write a definitive version of the notes.

[1] Ihara (1968) introduced the term "Shimura curve" for the "algebraic curves uniformized by automorphic functions attached to quaternion algebras over totally real fields, whose beautiful arithmetic properties have been discovered by Shimura (Annals 1967)." Langlands (1976) introduced the term "Shimura variety" for "certain varieties" studied "very deeply" by Shimura. His definition is that of Deligne 1971b.

[2] "Dans un petit nombre de cas, X/Γ peut s'interpréter comme l'ensemble des classes d'isomorphie des variétés abéliennes complexes, muni de quelques structures algébriques additionelles (polarisations, endomorphismes, structures sur les points d'ordre n)." Deligne 1971b

NOTATION AND CONVENTIONS

Throughout, k is a field. 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 k-vector space (or module) V is denoted by $V^\vee$. For a k-vector space V and a commutative k-algebra R, $V(R)$ denotes $V \otimes_k R$ (and similarly for $\mathbb{Z}$-modules). By a lattice, we always mean a full lattice. For example, a lattice in an $\mathbb{R}$-vector space V is a $\mathbb{Z}$-submodule Λ such that $\Lambda \otimes_{\mathbb{Z}} \mathbb{R} \simeq V$ — throughout $\simeq$ denotes a *canonical* isomorphism. The symbol k^{a} denotes an algebraic closure of the field k and k^{s} the separable closure of k in k^{a}. The transpose of a matrix C is denoted by C^t.

An algebraic group over a field k is a group scheme of finite type over k. As k is always of characteristic zero, such groups are smooth, and hence are not essentially different from the algebraic groups in Borel 1991 or Springer 1998. Let G be an algebraic group over a field k of characteristic zero. If G is connected or, more generally, if every connected component of G has a k-point, then $G(k)$ is dense in G for the Zariski topology (Milne 2017, 17.93). This implies that a connected algebraic subgroup of an algebraic group over k is determined by its k-points, and that a homomorphism from a connected algebraic group is determined by its action on the k-points.

Semisimple and reductive groups, whether algebraic or Lie, are required to be connected. A simple algebraic or Lie group is a semisimple group whose only proper normal subgroups are finite (sometimes such a group is said to be almost-simple). For example, SL_n is simple. For a torus T over k, $X^*(T)$ denotes the character group of $T_{k^{\mathrm{a}}}$. The derived group of a reductive group G is denoted by G^{der} (it is a semisimple group), and the adjoint group (quotient of G by its centre) is denoted by G^{ad}. Let $g \in G(k)$; then g acts on G by the inner automorphism $\mathrm{ad}(g) \overset{\text{def}}{=} (x \mapsto gxg^{-1})$ and hence on $\mathrm{Lie}(G)$ by an automorphism $\mathrm{Ad}(g)$. For more notation concerning reductive groups, see §5. For a finite extension of fields $L \supset F$, the algebraic group over F obtained by restriction of scalars from an algebraic group G over L is denoted by $(G)_{L/F}$.

A superscript $^+$ (resp. $^\circ$) denotes a connected component relative to a real topology (resp. a Zariski topology). For an algebraic group, it means the identity connected component. For example, $(\mathrm{O}_n)^\circ = \mathrm{SO}_n$, $(\mathrm{GL}_n)^\circ = \mathrm{GL}_n$, 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 Hausdorff.

Throughout, I use the notation standard in algebraic geometry, which sometimes conflicts with that 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 $\sigma: k \hookrightarrow K$ 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 that B classifies the elements of A modulo $\sim$ or that it classifies the $\sim$-classes of elements of A. The cardinality of a set S is denoted by $|S|$. Throughout, I write $A \backslash B \times C/D$ for the double coset space $A \backslash (B \times C)/D$ (apply $\times$ before $\backslash$ and $/$).

A functor $F: \mathsf{A} \to \mathsf{B}$ 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. An equivalence is a fully faithful functor $F\colon \mathsf{A} \to \mathsf{B}$ whose essential image is B.

REFERENCES

In addition to the references listed at the end, I refer to the following of my course notes (available at www.jmilne.org/math/).
AG: Algebraic Geometry, v6.02, 2017.
ANT: Algebraic Number Theory, v3.07, 2017.
CFT: Class Field Theory, v4.02, 2013.
MF: Modular Functions and Modular Forms, v1.31, 2017.

PREREQUISITES

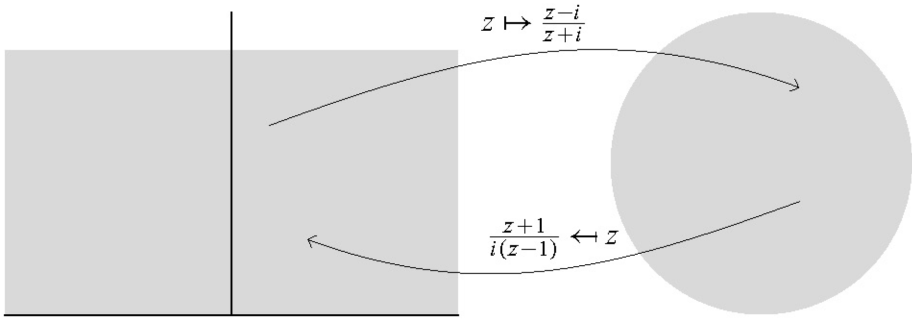
Beyond the mathematics that students usually acquire by the end of their first year of graduate work (a little complex analysis, topology, algebra, differential geometry,...), I assume some familiarity with algebraic number theory, algebraic geometry, algebraic groups, and elliptic modular curves.

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 Bin Du, Brian Conrad, F. Hörmann, and Shenhao Sun for alerting me to errors in earlier versions. The current version (2017), takes account of comments and corrections by Alexey Beshenov, Bas Edixhoven, Lucio Guerberoff, and Bhupendra Nath Tiwari. I especially thank Bjorn Poonen and the participants in his seminar (Yuzhou Gu, Naoki Imai, Jun Su, Isabel Vogt, Lynnelle Ye) for their careful reading of the original version and for their copious comments

1 Hermitian symmetric domains

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



Complex upper half plane, $\mathcal{H}_1$: $\Im(z) > 0$ Open unit disk, $\mathcal{D}_1$: $|z| < 1$

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

Brief review of real manifolds

A topological space M is ***locally euclidean at*** $p \in M$ if there exists an n such that p has an open neighbourhood homeomorphic to an open subset of $\mathbb{R}^n$. It is a ***manifold*** if it is locally euclidean at every point, Hausdorff, and admits a countable base for its open sets. A homeomorphism $\xi = (x^1, \ldots, x^n): U \to \mathbb{R}^n$ from an open subset of M onto an open subset of $\mathbb{R}^n$ is called a ***chart*** of M, and $x^1, \ldots, x^n$ are said to be ***local coordinates*** on U.

SMOOTH MANIFOLDS

We 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 $u_\alpha: U_\alpha \to \mathbb{R}^n$ of charts such that $M = \bigcup U_\alpha$ and the maps

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

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

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*** $\mathrm{Tgt}_p(M)$ to M at p is the $\mathbb{R}$-vector space of $\mathbb{R}$-derivations $X_p: \mathcal{O}_{M,p} \to \mathbb{R}$. If $x^1, \ldots, x^n$ are local coordinates at p, then $\left\{\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right\}$ is a basis for $\mathrm{Tgt}_p(M)$ and $\left\{dx^1, \ldots, dx^n\right\}$ is the dual basis. A smooth map $\alpha: M \to M'$ defines a linear map $d\alpha_p: \mathrm{Tgt}_p(M) \to \mathrm{Tgt}_{\alpha(p)}(M')$ for every $p \in M$.

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

A ***riemannian manifold*** is a smooth manifold M endowed with a ***riemannian metric***, i.e., a smooth 2-tensor field g such that, for all $p \in M$,

$$g_p\colon \mathrm{Tgt}_p(M) \times \mathrm{Tgt}_p(M) \to \mathbb{R}$$

is symmetric and positive-definite. In terms of local coordinates[3] $x^1, \ldots, x^n$ at p,

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

An isomorphism of riemannian manifolds is called an ***isometry***.

A ***real Lie group*** G is a smooth manifold endowed with a group structure defined by smooth maps $g_1, g_2 \mapsto g_1 g_2$, $g \mapsto g^{-1}$. According to a theorem of Lie, this is equivalent to the usual definition in which "smooth" is replaced by "real-analytic". A Lie group is ***adjoint*** if it is semisimple with trivial centre.

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

$$\begin{aligned}
&\varphi \text{ is symmetric} && \varphi(Ju, Jv) = \varphi(u,v), && (2)\\
&\psi \text{ is alternating} && \psi(Ju, Jv) = \psi(u,v), && (3)\\
&\psi(u,v) = -\varphi(u, Jv), && \varphi(u,v) = \psi(u, Jv). && (4)
\end{aligned}$$

Conversely, if φ satisfies (2), then the formulas (4) and (1) define a hermitian form,

$$(u|v) = \varphi(u,v) + i\varphi(u, Jv). \tag{5}$$

As $(u|u) = \varphi(u,u)$, the hermitian form $(\,|\,)$ is positive-definite if and only if φ is positive-definite. Note that φ is the bilinear form associated with the quadratic form $u \mapsto (u|u)\colon V \to \mathbb{R}$.

[3] In this situation, we usually write $dx^i dx^j$ for $dx^i \otimes dx^j$ — see Lee 1997, p. 24 for an explanation.

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 $u_\alpha: U_\alpha \to \mathbb{C}^n$ 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 $\alpha: M \to N$ of complex manifolds is ***analytic*** if it is a morphism 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 $\mathrm{Tgt}_p(M)$ (M regarded as a complex manifold) and $\mathrm{Tgt}_p(M^\infty)$ (M regarded as a smooth manifold) can be identified (as real vector spaces). Explicitly, complex local coordinates $z^1, \ldots, z^n$ at a point p of M define real local coordinates $x^1, \ldots, x^n, y^1, \ldots, y^n$ when we write $z^r = x^r + iy^r$. The real tangent space has basis $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial y^1}, \ldots, \frac{\partial}{\partial y^n}$ and the complex tangent space has basis $\frac{\partial}{\partial z^1}, \ldots, \frac{\partial}{\partial z^n}$. Under the natural identification of the two spaces we have $\frac{\partial}{\partial z^r} = \frac{1}{2}\left(\frac{\partial}{\partial x^r} - i\frac{\partial}{\partial y^r}\right)$.

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 $df_p: \mathrm{Tgt}_p(U) \to \mathrm{Tgt}_{f(p)}(\mathbb{C})$ 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 $df_p: \mathrm{Tgt}_p(U) \to \mathrm{Tgt}_{f(p)}(\mathbb{C}) = \mathbb{C}$ are $\mathbb{C}$-linear for all $p \in U$.

An ***almost-complex structure*** on a smooth manifold M is a smooth tensor field $J = (J_p)_{p \in M}$,

$$J_p: \mathrm{Tgt}_p(M) \to \mathrm{Tgt}_p(M),$$

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 $\alpha: M \to N$ of complex manifolds is holomorphic (analytic) if and only if the maps $d\alpha_p: \mathrm{Tgt}_p(M) \to \mathrm{Tgt}_{\alpha(p)}(N)$ 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 $\mathrm{Tgt}_p M$, which explains the name. A ***hermitian manifold*** (M,g) is a complex manifold M with a hermitian metric g, or, in other words, it is a riemannian manifold (M^∞, g) 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, i.e., for every pair of points p,q, there is an automorphism sending p to q. In particular, every point on the manifold then looks exactly like every other point.

SYMMETRIC SPACES

A manifold (riemannian, hermitian, ...) is ***symmetric*** if it is homogeneous and 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. By homogeneity, there is then a symmetry at every point.

The automorphism group of a riemannian manifold (M,g) 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.

ASIDE. Let (M,g) be a connected riemannian manifold. For each $p \in M$ and $v \in \mathrm{Tgt}_p(M)$ there is a unique maximal geodesic $\gamma: I \to M$ with $\gamma(0) = p$ and $\dot\gamma(0) = v$. Here I is an interval in $\mathbb{R}$. For each $p \in M$, there is a diffeomorphism defined on a neighbourhood of p (the ***geodesic symmetry at*** p) that sends $\gamma(t)$ to $\gamma(-t)$ for every geodesic γ with $\gamma(0) = p$. Geometrically, it is reflection along geodesics through p. If the geodesic symmetry at p is an isometry, then (M,g) is said to be ***locally symmetric at*** p. A connected riemannian manifold (M,g) is symmetric if and only if the geodesic symmetry at every point of M is an isometry and extends to an isometry of order 2 on the whole of M.

We let $\mathrm{Hol}(M)$ denote the group of automorphisms of a complex manifold M. The automorphism group of a hermitian manifold (M,g) 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 $\mathrm{Aut}(M^\infty)$). 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 $x+iy \in \mathcal{H}_1$, $x+iy = \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix} i$, 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 the geodesics (great circles) through a point p is a symmetry (this is equal to rotation through π about an axis through p and its polar opposite). These transformations leave the metric invariant.

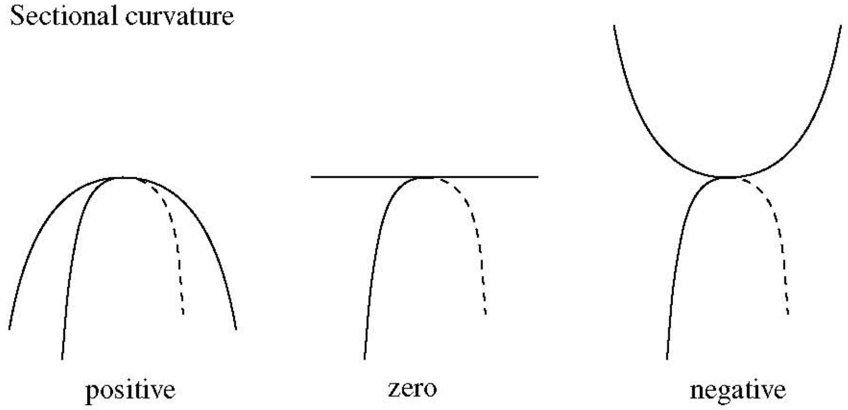
(c) Every quotient $\mathbb{C}/\Lambda$ of $\mathbb{C}$ by a discrete additive subgroup Λ becomes a hermitian symmetric space when endowed with the standard metric. The group of translations is transitive, and $z \mapsto -z$ is a symmetry at 0.

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 a chosen normal or away from it). 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.

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.

Sectional curvature



More generally, for a point p on a riemannian manifold M of any dimension, 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 $\mathrm{Tgt}_p M$. If the sectional curvature $K(p,E)$ is positive (resp. negative, resp. zero) for all p and E, then M is said to have positive (resp. negative, resp. zero) curvature.

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[4] (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 (Helgason 1978, VIII; Wolf 1984, 8.7):

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

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. Both M^- and M^+ are products of irreducible hermitian symmetric spaces, each having 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 $Z = X + iY$ with positive-definite imaginary part Y. 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 $\sum_{i=1}^{g} x_i y_{-i} - \sum_{i=1}^{g} x_{-i} y_i$:

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

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

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

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.

[4] Henri Cartan proved that the group of isometries of a bounded domain has a natural structure of a Lie group. Then his father Emil Cartan proved that the group of isometries of a symmetric bounded domain is semisimple (Borel 2001, IV 6). Myers and Steenrod proved that the group of isometries of every riemannian manifold is a Lie group.

THEOREM 1.3. *Every bounded domain has a canonical hermitian metric (called the Bergman metric).*[5] *Moreover, this metric has negative curvature.*

SKETCH OF PROOF. We ignore convergence questions. Initially, let D be any domain in $\mathbb{C}^n$. The holomorphic square-integrable functions $f: D \to \mathbb{C}$ form a Hilbert space $H(D)$ with inner product $(f|g) = \int_D f\bar{g}dv$. The first step is to prove that there is a unique (Bergman kernel) function $K: D \times D \to \mathbb{C}$ 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) $f(z) = \int_D K(z,\zeta)f(\zeta)dv(\zeta)$ for all $f \in H(D)$.

Let $(e_m)_{m\in\mathbb{N}}$ be a complete orthonormal set in $H(D)$, and let

$$K(z,\zeta) = \sum\nolimits_m e_m(z)\cdot\overline{e_m(\zeta)}.$$

Obviously $K(z,\zeta) = \overline{K(\zeta,z)}$, and

$$f = \sum\nolimits_m (f|e_m)e_m = \int K(\cdot,\zeta)f(\zeta)dv(\zeta)$$

(actual equality, not almost-everywhere equality, because the functions are holomorphic). Therefore K satisfies (b) and (c), and we are ignoring (a). Let k be a second function satisfying (a), (b), and (c). Then

$$k(z,\zeta) = \int_D K(z,t)k(t,\zeta)dv(t) = \overline{\int_D k(\zeta,t)\overline{K(z,t)}dv(t)} = K(z,\zeta),$$

which proves the uniqueness.

Now assume that 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

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

is 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 of $\mathrm{Hol}(D)$. 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 $I_g - \bar{Z}^t Z$ 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.

[5] After Stefan Bergmann. When he moved to the United States in 1939, he dropped the second n from his name.

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, the stablizer K_p of p in $\mathrm{Hol}(M)^+$ is compact, and $\mathrm{Hol}(M)^+/K_p \simeq M^\infty$.

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 Lemma 1.5. □

Let H be a connected real Lie group with Lie algebra $\mathfrak{h}$. There need not be an algebraic group G over $\mathbb{R}$ such that $G(\mathbb{R})^+ = H$. 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 which is algebraic because SL_2 is simply connected[6] as an algebraic group. However, if H admits 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 $\mathfrak{gl}(V)$) (Borel 1991, 7.9). When H is semisimple, $[\mathfrak{h}, \mathfrak{h}] = \mathfrak{h}$, and so $\mathrm{Lie}(G) = \mathfrak{h}$. This implies that $G(\mathbb{R})^+ = H$ (inside $\mathrm{GL}(V)$).

PROPOSITION 1.7. *Let (M,g) be a hermitian symmetric domain, and let $\mathfrak{h}$ denote the Lie algebra of $\mathrm{Hol}(M)^+$. 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 stabilizer in $G(\mathbb{R})$ of M.

[6] A connected algebraic group G in characteristic zero is ***simply connected*** if every isogeny (surjective homomorphism with finite kernel) $G' \to G$ is an isomorphism. Every semisimple algebraic group admits an essentially unique isogeny $\tilde{G} \to G$ with $\tilde{G}$ connected and simply connected.

PROOF. The Lie group $\mathrm{Hol}(M)^+$ is adjoint (because $\mathrm{Is}(M,g)^+$ is adjoint), and so the adjoint representation realizes it as a subgroup of $\mathrm{GL}(\mathfrak{h})$. The first statement now follows from the above discussion, and the second follows from Satake 1980, 8.5. □

The algebraic group G in the proposition is adjoint (in particular, semisimple) and $G(\mathbb{R})$ is not compact.

EXAMPLE 1.8. The map $z \mapsto \bar{z}^{-1}$ is an antiholomorphic isometry of $\mathcal{H}_1$, and every isometry of $\mathcal{H}_1$ is either holomorphic or differs from $z \mapsto \bar{z}^{-1}$ 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 $u_p: U_1 \to \mathrm{Hol}(D)$

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 $u_p: U_1 \to \mathrm{Hol}(D)$ such that $u_p(z)$ fixes p and acts on $\mathrm{Tgt}_p(D)$ as multiplication by z.*

EXAMPLE 1.10. Let $p = i \in \mathcal{H}_1$, and let $h: \mathbb{C}^\times \to \mathrm{SL}_2(\mathbb{R})$ be the homomorphism sending $z = a + ib$ to $\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$. Then $h(z)$ acts on the tangent space $\mathrm{Tgt}_i(\mathcal{H}_1)$ as multiplication by $z/\bar{z}$, because $\frac{d}{dz}\left(\frac{az+b}{-bz+a}\right)\Big|_i = \frac{a^2+b^2}{(a-bi)^2}$. 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 $\mathrm{Tgt}_i \mathcal{H}_1$ as multiplication by z.

Because of the importance of the theorem, I explain the proof. A riemannian manifold is ***geodesically complete*** (or just ***complete***) if every maximal geodesic is defined on the whole of $\mathbb{R}$.

PROPOSITION 1.11. *Let (M,g) be a symmetric space.*

(a) *Let $p \in M$. The symmetry s_p at p acts as -1 on $\mathrm{Tgt}_p(M)$, and it sends $\gamma(t)$ to $\gamma(-t)$ for every geodesic γ with $\gamma(0) = p$.*

(b) *The pair (M,g) is geodesically complete.*

PROOF. (a) Because $s_p^2 = 1$, $(ds_p)^2 = 1$, and so ds_p acts semisimply on $\mathrm{Tgt}_p M$ with eigenvalues ± 1. Let X be a tangent vector at p and $\gamma: I \to M$ the (unique) maximal geodesic with $\gamma(0) = p$ and $\dot{\gamma}(0) = X$. If $(ds_p)(X) = X$, then $s_p \circ \gamma$ is a geodesic also with these properties, and so p is not an isolated fixed point of s_p. Therefore only -1 occurs as an eigenvalue of ds_p. If $(ds_p)(X) = -X$, then $s_p \circ \gamma$ and $t \mapsto \gamma(-t)$ are geodesics through p with initial tangent vector $-X$, and so they are equal.

(b) Let $\gamma: I \to M$ be a maximal geodesic. If $I \neq \mathbb{R}$, then a symmetry $s_{\gamma(t_0)}$ with t_0 an end point of I maps the geodesic γ to another geodesic through $\gamma(t_0)$ extending γ. This contradicts the maximality of γ. (See Boothby 1975, VII 8.4.) □

By a ***canonical tensor*** on a symmetric space (M,g), we mean a tensor fixed by every isometry of (M,g). Every tensor "canonically derived" from g will be canonical. On a riemannian manifold, there is a well-defined (riemannian) connection, and hence the notion of "parallel translation" of tangent vectors along geodesics.

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 it is zero. This implies that parallel translation of 2-dimensional subspaces of tangent spaces does not change the sectional curvature. □

We shall need the notion of the ***exponential map*** at a point p of a riemannian manifold (M,g). For $v \in \mathrm{Tgt}_p(M)$, let $\gamma_v: I_v \to M$ denote the maximal geodesic with $\gamma_v(p) = p$ and $\dot{\gamma}(0) = v$. Let D_p be the set of $v \in \mathrm{Tgt}_p(M)$ such that I_v contains 1. Then there is a unique map $\exp_p: D_p \to M$ such that

$$\exp_p(tv) = \gamma_v(t)$$

whenever $tv \in D_p$. Moreover, $\exp_p$ is smooth on an open neighbourhood U of 0 in D_p. The map on tangent spaces $\left(d \exp_p\right)_0: \mathrm{Tgt}_0(U) \to \mathrm{Tgt}_p(M)$ is the canonical isomorphism. If M is geodesically complete, then $\exp_0$ is defined on the whole of $\mathrm{Tgt}_p(M)$. See Lee 1997, Chapter 5.

PROPOSITION 1.13. *Let (M,g) and (M',g') be riemannian manifolds in which parallel translation of 2-dimensional subspaces of tangent spaces does not change the sectional curvature. Let $a: \mathrm{Tgt}_p(M) \to \mathrm{Tgt}_{p'}(M')$ be a linear isometry such that $K(p,E) = K(p',aE)$ for every 2-dimensional subspace $E \subset \mathrm{Tgt}_p(M)$. 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 riemannian metrics in terms of normal geodesic coordinates. See Wolf 1984, 2.3.7. □

PROPOSITION 1.14. *Let (M,g), (M',g'), and $a: \mathrm{Tgt}_p(M) \to \mathrm{Tgt}_{p'}(M')$ be as in 1.13. If M and M' are complete, connected, and simply connected, then there is a unique isometry $\alpha: M \to M'$ such that $\alpha(p) = p'$ and $(d\alpha)_p = a$.*

PROOF. The conditions imply that the locally-defined isometry in 1.13 extends globally — see Wolf 1984, 2.3.12. □

We now prove Theorem 1.9. Let $p \in D$. Each complex number z with $|z| = 1$ defines an automorphism of the vector space $\mathrm{Tgt}_p(D)$ preserving g_p, and one checks that it also preserves sectional curvatures. According to Propositions 1.11, 1.12, and 1.14, there exists a unique isometry $u_p(z): D \to D$ fixing p and acting as multiplication by z on $\mathrm{Tgt}_p(D)$. 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 $\mathrm{Tgt}_p(D)$, and so it equals $u_p(zz')$.

Cartan involutions

Let G be a connected algebraic group over $\mathbb{R}$, and let $g \mapsto \bar{g}$ 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

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

is compact.

EXAMPLE 1.15. Let $G = \mathrm{SL}_2$, and let $\theta = \mathrm{ad}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Then

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

for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{C})$, and so

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

The map $\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \mapsto (a,b)$ identifies $\mathrm{SU}_2(\mathbb{R})$ with a closed bounded set in $\mathbb{C}^2$, and so it is compact. Therefore θ is a Cartan involution of SL_2.

THEOREM 1.16. *Let G be a connected algebraic group over k. There exists a Cartan involution of G if and only if G is reductive, in which case any two are conjugate by an element of $G(\mathbb{R})$ (i.e., they differ by $\mathrm{ad}\, g$ for some $g \in G(\mathbb{R})$).*

PROOF. See Satake 1980, I 4.3. □

EXAMPLE 1.17. Let G be a connected algebraic group over $\mathbb{R}$. We say that G is ***compact*** if $G(\mathbb{R})$ is compact.

(a) If the identity map on G is a Cartan involution, then $G(\mathbb{R})$ is compact. Conversely, if $G(\mathbb{R})$ is compact, then the identity map is a Cartan involution, and it is the only Cartan involution because of the second part of the theorem.

(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 implies 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 $g \mapsto \theta(\bar{g})$. Therefore θ is Cartan if and only if $G^{(\theta)}$ is compact. All compact real forms of $G_{\mathbb{C}}$ arise in this way from a Cartan involution of G.

PROPOSITION 1.18. *Let G be a connected algebraic group over $\mathbb{R}$. If G 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 $\rho: G \to \mathrm{GL}(V)$ 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. □

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 $\varphi(gu, \bar{g}v) = \varphi(u,v)$, $g \in G(\mathbb{C})$, $u, v \in V$.)

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 and $\mathrm{ad}(C) = \mathrm{ad}(C^{-1})$). 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. We first remark that an $\mathbb{R}$-bilinear form φ on a real vector space V extends to a sesquilinear form φ' on $V(\mathbb{C})$,

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

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

Let $\rho: G \to \mathrm{GL}(V)$ 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

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

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

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

This can be rewritten as

$$\varphi'_C(gu, ((\mathrm{ad}\, C)\bar{g})v) = \varphi'_C(u,v),$$

where $\varphi'_C = (\varphi_C)'$. This last equation 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). Thus $\mathrm{ad}\, C$ 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. □

Representations of U_1

Let T be a torus over a field k. If T is split, then every representation $\rho: T \to \mathrm{GL}_V$ is diagonalizable. This means that $V = \bigoplus_{\chi \in X^*(T)} V_\chi$, where V_χ is the subspace on which T acts through the character χ,

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

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

Now suppose that T splits only over a Galois extension K of k. Let V be a k-vector space, and let ρ be a representation of T_K on $K \otimes V$. Then $K \otimes V = \bigoplus_{\chi \in X^*(T)} V_\chi$, and a direct calculation shows that ρ is fixed by an element σ of $\mathrm{Gal}(K/k)$ if and only if $\sigma V_\chi = V_{\sigma\chi}$ for all χ. Therefore ρ is defined over k if and only if

$$\sigma(V_\chi) = V_{\sigma\chi}, \quad \text{all } \sigma \in \mathrm{Gal}(K/k), \quad \chi \in X^*(T). \tag{12}$$

It follows that to give a representation of T on a k-vector space V amounts to giving a gradation $K \otimes V = \bigoplus_{\chi \in X^*(T)} V_\chi$ of $K \otimes V$ for which (12) holds (Milne 2017, 12.30).

When we regard U_1 as a real algebraic torus, its characters are $z \mapsto z^n$, $n \in \mathbb{Z}$. Thus $X^*(U_1) \simeq \mathbb{Z}$, and complex conjugation acts on it as multiplication by -1. Therefore a representation of U_1 on a real vector space V is a gradation $V(\mathbb{C}) = \bigoplus_{n \in \mathbb{Z}} V^n$ such that $V(\mathbb{C})^{-n} = \overline{V(\mathbb{C})^n}$ (complex conjugate) for all n. 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.[7] The natural map

$$V/V^0 \to V(\mathbb{C})/\textstyle\bigoplus_{n \leq 0} V(\mathbb{C})^n \simeq \bigoplus_{n>0} V(\mathbb{C})^n \tag{13}$$

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

These representations are irreducible, and no two are isomorphic.

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 real Lie groups is algebraic. It follows that the homomorphisms

$$u_p: U_1 \to \mathrm{Hol}(D)^+ \overset{1.7}{\simeq} G(\mathbb{R})^+$$

in Theorem 1.9 are 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 $u_p: U_1 \to G$ attached to a point p of D has the following properties:*

[7] We are using the following statement. Let K be a Galois extension of k with Galois group Γ, and let V_0 be a vector space over k. The rule $\sigma(a \otimes v) = \sigma a \otimes v$ defines a semilinear action of Γ on $V = K \otimes V_0$. A subspace W of V is of the form KW_0 with W_0 a subspace of V_0 if and only if it is stable under the action of Γ, in which case the map $a \otimes w \mapsto aw: K \cap (V_0 \cap W) \to W$ is an isomorphism.

(a) *only the characters* $z, 1, z^{-1}$ *occur in the representation of* U_1 *on* $\mathrm{Lie}(G)_{\mathbb{C}}$ *defined by* $\mathrm{Ad} \circ u_p$*;*

(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. According to Proposition 1.6,

$$G(\mathbb{R})^+/K_p \simeq D \quad \text{(isomorphism of smooth manifolds)},$$

where K_p is the subgroup fixing p. For $z \in U_1$, the action of $u_p(z)$ on $G(\mathbb{R})^+$ by conjugation preserves K_p and corresponds to the obvious action of $u_p(z)$ on D. On passing to the tangent spaces, we obtain an isomorphism of real vector spaces

$$\mathrm{Lie}(G)/\mathrm{Lie}(K_p) \simeq \mathrm{Tgt}_p D.$$

By definition, $u_p(z)$ acts on $\mathrm{Tgt}_p(D)$ as multiplication by z. Statement (a) follows from this because $u_p(z)$ acts trivially on $\mathrm{Lie}(K_p)$.[8]

The symmetry s_p at p and $u_p(-1)$ both fix p and act as -1 on $\mathrm{Tgt}_p D$ (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 $\{g \in G(\mathbb{C}) \mid g = u(-1) \cdot \bar{g} \cdot u(-1)^{-1}\}$, which, according to (b), is compact. As K_u is closed, it also is compact. The equality $D = \left(G(\mathbb{R})^+/K_u\right) \cdot u$ endows D with the structure of smooth (even real-analytic) manifold. For this structure, the tangent space to D at u,

$$\mathrm{Tgt}_u(D) \simeq \mathrm{Lie}(G)/\mathrm{Lie}(K_u) \simeq \mathrm{Lie}(G)/\mathrm{Lie}(G)^0,$$

which, because of (a), can be identified with the subspace of $\mathrm{Lie}(G)_{\mathbb{C}}$ on which $u(z)$ acts as z (see (13)). This endows $\mathrm{Tgt}_u D$ 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 be integrable (Wolf 1984, 8.7.9). The action of K_u on D defines an action of it on $\mathrm{Tgt}_u D$. Because K_u is compact, there is a K_u-invariant positive-definite form on $\mathrm{Tgt}_u D$ (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)). □

[8]In fact, $u_p(z)$ acts trivially on K_p: if $k \in K_p$, then Proposition 1.14 shows that $u_p(z) \cdot k \cdot u_p(z)^{-1}$ equals k because they both fix p and act as $(dk)_p$ on $\mathrm{Tgt}_p(D)$.

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

Replacing a point of the domain with a second point replaces the homomorphism u with a conjugate. Therefore we get a natural one-to-one correspondence between the isomorphism classes of hermitian symmetric domains and of pairs $(G,[u])$ consisting of a real adjoint Lie group and a conjugacy class of homomorphisms $u: U_1 \to G(\mathbb{R})$ satisfying (a), (b), (c) of 1.21.

EXAMPLE 1.23. Let $u: U_1 \to \mathrm{PSL}_2(\mathbb{R})$ 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 $\operatorname{ad} u(-1)$ is a Cartan involution of SL_2, hence also of PSL_2. The corresponding hermitian symmetric domain is $\mathcal{H}_1$.

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 G is an inner form of its compact form,[9] 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 H over $\mathbb{R}$ and a conjugacy class of homomorphisms $u: U_1 \to H$ 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 $g \mapsto \bar{g}$ denote complex conjugation on $G(\mathbb{C})$ relative to a maximal compact subgroup of G containing $\mu(U_1)$.[10] There is a real form H of G such that complex conjugation on $H(\mathbb{C}) = G(\mathbb{C})$ is $g \mapsto \mu(-1)\cdot\bar{g}\cdot\mu(-1)^{-1}$, and $u \overset{\text{def}}{=} \mu|U_1$ 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 over $\mathbb{C}$. 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*** $\tilde{\alpha} = \sum n_i\alpha_i$ with the property that, for every other root $\sum m_i\alpha_i$, the coefficient $n_i \geq m_i$ for 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

[9] If $G_{\mathbb{C}}$ is not simple, say, $G_{\mathbb{C}} = G_1 \times G_2$, then $G = \mathrm{Res}_{\mathbb{C}/\mathbb{R}}(G_1)$ and any inner form of G is also the restriction of scalars of a $\mathbb{C}$-group; but such a group cannot be compact (look at a subtorus).

[10] Take the complex multiplication defined by any real form G_0 of G, and modify it by a suitable Cartan involution G_0, which exists because G_0 is reductive (1.16).

the Weyl group acts simply transitively on the Weyl chambers, 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 $\langle \alpha, \mu \rangle \in \{1, 0, -1\}$ for all roots α.[11] 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. We have proved the following statement.

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 tables in Bourbaki 1981 or Helgason 1978, p. 477. One obtains the following table:

type	$\tilde{\alpha}$	special roots	#
A_n	$\alpha_1 + \alpha_2 + \cdots + \alpha_n$	$\alpha_1, \ldots, \alpha_n$	n
B_n	$\alpha_1 + 2\alpha_2 + \cdots + 2\alpha_n$	α_1	1
C_n	$2\alpha_1 + \cdots + 2\alpha_{n-1} + \alpha_n$	α_n	1
D_n	$\alpha_1 + 2\alpha_2 + \cdots + 2\alpha_{n-2} + \alpha_{n-1} + \alpha_n$	$\alpha_1, \alpha_{n-1}, \alpha_n$	3
E_6	$\alpha_1 + 2\alpha_2 + 2\alpha_3 + 3\alpha_4 + 2\alpha_5 + \alpha_6$	α_1, α_6	2
E_7	$2\alpha_1 + 2\alpha_2 + 3\alpha_3 + 4\alpha_4 + 3\alpha_5 + 2\alpha_6 + \alpha_7$	α_7	1
E_8, F_4, G_2		none	0

Mnemonic: the number of special simple roots is one less than the connection index of the root system.

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, nontrivial compact groups do not arise in this way, and PGL_n does so only for $n = 2$.[12] Starting from the above table, it is possible to list them. For the classical groups, we do this in Addendum B. There is a complete list in Lan 2017, §3,

ASIDE. Why do we consider hermitian *symmetric* spaces rather than *homogeneous* spaces? The short answer is that the connection between symmetric spaces and semisimple groups was recognized early (by Elie Cartan), and the subject matured quickly. Until 1957, no examples of nonsymmetric bounded homogeneous domains in $\mathbb{C}^n$ were known. As Borel writes (1998)

> The famous Math. Zeitschrift papers by Hermann Weyl (1925–26) mark the beginning of the global theory of semisimple Lie groups. They had right away a considerable impact on Elie Cartan. At the time Cartan was determining locally symmetric spaces, via their holonomy groups, and had discovered with surprise that this was equivalent to one he had solved about twelve years earlier: the classification of real forms of complex semisimple Lie algebras. Under the influence of Weyl's papers, he soon cast this work in a global framework and built up a beautiful theory.

I do not know how much of the theory in this article extends to quotients of bounded homogeneous domains.

[11]The μ with this property are sometimes said to be ***minuscule*** (cf. Bourbaki 1981, pp. 226–227).

[12]Suppose that $G = \mathrm{PGL}_n$, $n \geq 2$, arises from a hermitian symmetric domain, and let $u = u_p$ be as in Theorem 1.21. The theorem and its proof show that $u(U_1)$ is contained in the centre of the maximal compact subgroup K_u of $G(\mathbb{R})^+$ defined by the Cartan involution $\mathrm{ad}(u(-1))$. Therefore the centre of K_u is nondiscrete, but the centres of the maximal compact subgroups of $\mathrm{PGL}_n(\mathbb{R})^+$(they are all conjugate) are discrete if $n > 2$.

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. The present account follows Deligne 1973a and Deligne 1979. For a history of symmetric spaces, see Chapter IV of Borel 2001.

2 Hodge structures and their classifying spaces

In this section, following Deligne, we interpret hermitian symmetric domains as parameter spaces for Hodge structures. Later, this will allow us to realize certain quotients of bounded symmetric domains as moduli varieties of abelian varieties or, more generally, abelian motives.

Reductive groups and tensors

Let G be a reductive group over a field k of characteristic zero, and let $\rho: G \to \mathrm{GL}(V)$ 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

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

— this is essentially the same as an element of $(V^{\otimes r})^\vee$. For an r-tensor t, the condition

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

on g defines an algebraic 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*** $t \in T$.

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. According to a theorem of Chevalley (Milne 2017, 4.27), G is the stabilizer in $\mathrm{GL}(V)$ of a one-dimensional subspace L of some representation W of $\mathrm{GL}(V)$. Because representations of reductive groups in characteristic zero are semisimple (ibid. 22.42), $W = W' \oplus L$ for some subspace W' of W stable under $\mathrm{GL}(V)$. Now G is the subgroup of $\mathrm{GL}(V)$ fixing any nonzero element t of $L \otimes L^\vee$ in $W \otimes W^\vee$. The representation $W \otimes W^\vee$ of $\mathrm{GL}(V)$ can be realized as a subrepresentation of a sum of representations $V^{\otimes r} \otimes V^{\vee\otimes s}$ (ibid. 4.14), and hence of a sum of representations $(V^{\otimes r})^\vee$. The image of t in $\bigoplus_r (V^{\otimes r})^\vee$ is a sum $t = \sum_r t_r$ with $t_r \in \left(V^{\otimes r}\right)^\vee$, and we can take T to be the set of nonzero t_r. □

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

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

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 $W \mapsto P(W): G_d(V) \to \mathbb{P}^{\binom{n}{d}-1}$ identifies $G_d(V)$ with a closed subvariety of $\mathbb{P}^{\binom{n}{d}-1}$ (AG, 6.29). A coordinate-free description of this map is given by

$$W \mapsto \bigwedge^d W : G_d(V) \to \mathbb{P}(\bigwedge^d V). \tag{14}$$

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,

$$G_d(V)_S \simeq \mathrm{Hom}(W_0, S). \tag{15}$$

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 is

$$\mathrm{Tgt}_{W_0}(G_d(V)) \simeq \mathrm{Hom}(W_0,S) \simeq \mathrm{Hom}(W_0, V/W_0). \tag{16}$$

FLAG VARIETIES

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

$$F: \quad V \supset V^1 \supset \cdots \supset V^r \supset 0 \tag{17}$$

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

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

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

$$\varphi^i : V^i \to V/V^i, \quad 1 \le i \le r, \tag{18}$$

satisfying the compatibility condition

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

ASIDE 2.3. A basis $e_1, \ldots, e_n$ for V is ***adapted to*** the flag F if it contains a basis $e_1, \ldots, e_{d_i}$ 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

$$g \mapsto gF : \mathrm{GL}(V)/P(F) \to G_{\mathbf{d}}(V)$$

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

Hodge structures

Hodge showed that the rational cohomology groups of a nonsingular projective algebraic over $\mathbb{C}$ carry an additional structure, now called a Hodge structure.

DEFINITION

For a real vector space V, complex conjugation on $V(\mathbb{C}) \overset{\text{def}}{=} \mathbb{C} \otimes_{\mathbb{R}} V$ is defined by

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

An $\mathbb{R}$-basis $e_1, \ldots, e_m$ 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 integer n, the subspace $\bigoplus_{p+q=n} V^{p,q}$ of $V(\mathbb{C})$ is stable under complex conjugation, and so it is defined over $\mathbb{R}$ (footnote 7, p. 17), i.e., there is a subspace V_n of V such that

$$V_n(\mathbb{C}) = \bigoplus_{p+q=n} V^{p,q}.$$

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 $\mathbb{C}$, and write $H^n(X,-)$ for $H^n(X(\mathbb{C}),-)$. Then $H^n(X,\mathbb{Q}) \otimes \mathbb{C} \simeq H^n(X,\mathbb{C})$, which can be computed using de Rham cohomology. This means that $H^n(X,\mathbb{C})$ is the nth cohomology group of the complex

$$\Omega^0(M) \to \cdots \to \Omega^m(M) \to \Omega^{m+1}(M) \to \cdots,$$

where $\Omega^m(M)$ is the space of $\mathbb{C}$-valued differential m-forms on the real manifold $M = X(\mathbb{C})$. Such a form is said to be of type (p,q) if it is everywhere locally a sum of terms of the following type:

$$f \cdot dz_1 \wedge \cdots \wedge dz_p \wedge d\bar{z}_1 \wedge \cdots \wedge d\bar{z}_q.$$

This gives a decomposition of each space $\Omega^m(M)$, which Hodge showed persists when we pass to the cohomology. In this way, we get a Hodge decomposition

$$H^n(X,\mathbb{C}) = \bigoplus\nolimits_{p+q=n} H^{p,q}$$

with $H^{p,q} \simeq H^q(X,\Omega^p)$. See Voisin 2002, 6.1.3.

EXAMPLE 2.6. We let $\mathbb{Q}(m)$ denote the (unique) Hodge structure of weight $-2m$ with underlying vector space $(2\pi i)^m\mathbb{Q}$. Thus, $(\mathbb{Q}(m))(\mathbb{C}) = \mathbb{Q}(m)^{-m,-m}$. We define $\mathbb{Z}(m)$ and $\mathbb{R}(m)$ similarly.[13]

THE HODGE FILTRATION

The ***Hodge filtration*** associated with a Hodge structure of weight n is

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

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

$$V^{p,q} = F^p \cap \overline{F^q}. \tag{19}$$

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).

[13]In the original version, I took the underlying vector space of $\mathbb{Q}(m)$ to be $\mathbb{Q}$. Adding the factor $(2\pi i)^m$ makes the entire theory invariant under a change of the choice of $i = \sqrt{-1}$ in $\mathbb{C}$. In the rest of the article, we sometimes include the $2\pi i$ and sometimes not.

Hodge structures as representations of $\mathbb{S}$

Let $\mathbb{S}$ be the algebraic torus over $\mathbb{R}$ obtained from $\mathbb{G}_m$ over $\mathbb{C}$ by restriction of the scalars — it is sometimes called the Deligne torus. Thus

$$\mathbb{S}(\mathbb{R}) = \mathbb{C}^\times \quad \text{and} \quad \mathbb{S}_\mathbb{C} \simeq \mathbb{G}_m \times \mathbb{G}_m.$$

We fix the second isomorphism so that the map $\mathbb{S}(\mathbb{R}) \to \mathbb{S}(\mathbb{C})$ induced by the inclusion $\mathbb{R} \hookrightarrow \mathbb{C}$ is $z \mapsto (z, \bar{z})$. Then $\mathbb{S}(\mathbb{C}) \simeq \mathbb{C}^\times \times \mathbb{C}^\times$ with complex conjugation acting by the rule $\overline{(z_1, z_2)} = (\overline{z_2}, \overline{z_1})$. The ***weight homomorphism***

$$w\colon \mathbb{G}_m \to \mathbb{S}$$

is the map such that $w(\mathbb{R})\colon \mathbb{G}_m(\mathbb{R}) \to \mathbb{S}(\mathbb{R})$ is $r \mapsto r^{-1}\colon \mathbb{R}^\times \to \mathbb{C}^\times$.

The characters of $\mathbb{S}_\mathbb{C}$ are the homomorphisms $(z_1, z_2) \mapsto z_1^r z_2^s$ with $(r, s) \in \mathbb{Z} \times \mathbb{Z}$. Thus, $X^*(\mathbb{S}) \simeq \mathbb{Z} \times \mathbb{Z}$ 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 $\mathbb{Z} \times \mathbb{Z}$-gradation of $V(\mathbb{C})$ such that $\overline{V^{p,q}} = V^{q,p}$ for all p, q (see p. 17). Thus, to give a representation of $\mathbb{S}$ on a real vector space V is the same as giving 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

$$h_\mathbb{C}(z_1, z_2)v = z_1^{-p} z_2^{-q} v \text{ for } v \in V^{p,q}. \tag{20}$$

In other words,

$$h(z)v = z^{-p} \bar{z}^{-q} v \text{ for } v \in V^{p,q}. \tag{21}$$

Note the minus signs! The associated weight decomposition has

$$V_n = \{v \in V \mid w_h(r)v = r^n v\}, \quad w_h = h \circ w. \tag{22}$$

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

$$\mu_h(z) = h_\mathbb{C}(z, 1). \tag{23}$$

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 V is a homomorphism $h\colon \mathbb{C} \to \mathrm{End}_\mathbb{R}(V)$ of $\mathbb{R}$-algebras. The restriction of this to a homomorphism $\mathbb{C}^\times \to \mathrm{GL}(V)$ is a Hodge structure of type $(-1, 0), (0, -1)$ whose associated complex structure (see 2.4) is that defined by h.[14]

Example 2.9. The Hodge structure $\mathbb{Q}(m)$ corresponds to the homomorphism

$$h\colon \mathbb{S} \to \mathbb{G}_{m\mathbb{R}}, \quad h(z) = (z\bar{z})^m.$$

[14]This partly explains the signs in (20); see also Deligne 1979, 1.1.6. Following Deligne 1973b, 8.12, and Deligne 1979, 1.1.1.1, the convention $h_\mathbb{C}(z_1, z_2)v^{p,q} = z_1^{-p} z_2^{-q} v^{p,q}$ is standard in the theory of Shimura varieties. Following Deligne 1971a, 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 common in Hodge theory (e.g., Voisin 2002, Chap. 6, Exer. 1).

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 $(V,(V^{-1,0},V^{0,-1})) \rightsquigarrow (V,C)$ 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 footnote 7, p. 17). Note that

$$V^{00} = \operatorname{Ker}(V \to V(\mathbb{C})/F^0). \tag{24}$$

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 $m+n$:

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

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. For example, a morphism of rational Hodge structures is a linear map $V \to W$ of $\mathbb{Q}$-vector spaces such that $V(\mathbb{R}) \to W(\mathbb{R})$ is a morphism 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

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

or if

$$\sum p_i \neq \sum q_i \Rightarrow t_{\mathbb{C}}(v_1^{p_1,q_1}, v_2^{p_2,q_2}, \ldots) = 0, \quad v_i^{p_i,q_i} \in V^{p_i,q_i}. \tag{25}$$

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 $\psi: V \times V \to \mathbb{R}(-n)$ such that $\psi_C(u,v) \overset{\text{def}}{=} (2\pi i)^n \psi(u,Cv)$ is symmetric and positive-definite. Then ψ is symmetric or alternating according as n is even or odd, because

$$\begin{aligned}\psi(v,u) = \psi(Cv,Cu) &= (2\pi i)^{-n}\psi_C(Cv,u)\\ &= (2\pi i)^{-n}\psi_C(u,Cv) = \psi(u,C^2 v) = \psi(u,(-1)^n v = (-1)^n \psi(u,v).\end{aligned}$$

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 $\psi: V \times V \to R$ such that $\psi_{\mathbb{R}}$ is a polarization of $(V(\mathbb{R}),h)$. A rational or real Hodge structure is ***polarizable*** if it admits a polarization on each component of the weight decomposition.

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 $\psi: V \times V \to 2\pi i R = R(1)$ such that, for $u,v \in V(\mathbb{R})$,

$$\begin{aligned}\psi_{\mathbb{R}}(Ju,Jv) &= \psi_{\mathbb{R}}(u,v), \text{ and}\\ \frac{1}{2\pi i}\psi_{\mathbb{R}}(u,Ju) &> 0 \text{ if } u \neq 0.\end{aligned}$$

Then the form $u,v \mapsto \psi_{\mathbb{R}}(u,Jv)$ 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 the Hodge structure $H^n(X,\mathbb{Q})$ for each n (Hodge index theorem; Voisin 2002, 6.3.2).

Variations of Hodge structures

> Consider a morphism $\pi: V \to S$ of nonsingular algebraic varieties over $\mathbb{C}$ whose fibres V_s, $s \in S$, are nonsingular projective varieties. The vector spaces $H^n(V_s,\mathbb{Q})$ form a local system of $\mathbb{Q}$-vector spaces on the topological space $S(\mathbb{C})$, and each space $H^n(V_s,\mathbb{Q})$ carries a Hodge structure. The Hodge decompositions on the spaces $H^n(V_s,\mathbb{Q})$ vary continuously in $s \in S(\mathbb{C})$, the Hodge filtrations vary holomorphically in $s \in S(\mathbb{C})$, and Griffiths showed that the system satisfies a certain (Griffiths) transversality condition. A system satisfying these conditions on a complex manifold is called a variation of Hodge structures.

In this section, we explain Deligne's realization of hermitian symmetric domains as parameter spaces for variations of Hodge structures. Because hermitian symmetric domains are simply connected, we need only consider variations in which the underlying local system is constant.

Let S be a connected complex manifold and V a real vector space. 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

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

into the Grassmannian is continuous.[15]

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

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

is holomorphic. Here $\mathbf{d} = (\ldots, d(p), \ldots)$, where $d(p) = \dim F_s^p V = \sum_{r \geq p} d(r,q)$.

For a holomorphic family of Hodge structures, the differential of φ at s is a $\mathbb{C}$-linear map

$$d\varphi_s : \mathrm{Tgt}_s S \to \mathrm{Tgt}_{F_s^\bullet}(G_{\mathbf{d}}(V(\mathbb{C}))) \overset{(18)}{\subset} \bigoplus\nolimits_p \mathrm{Hom}(F_s^p, V(\mathbb{C})/F_s^p).$$

We say that the family satisfies ***Griffiths transversality*** if the image of $d\varphi_s$ is contained in

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

for all s. When the family satisfies Griffiths transversality, we call it a ***variation of Hodge structures***.

A variation of Hodge structures on a nonconnected complex manifold is a variation of Hodge structures on each conneced component of the manifold.

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

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

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 subset of $\displaystyle\prod_{d(p,q)\neq 0} G_{d(p,q)}(V(\mathbb{C}))$.

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

(a) *The space 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.*[16]

(c) *Every irreducible hermitian symmetric domain is of the form S^+ for a suitable choice of V, d, and T.*

---

[15] Earlier we defined Grassmannians of a complex vector space to be algebraic varieties over $\mathbb{C}$. In this section, it is more convenient to regard them as complex manifolds.

[16] The converse is false: S^+ may be a hermitian symmetric domain without $(h_s)_s$ being a variation of Hodge structures. For example, let $V = \mathbb{R}^2$ with the standard alternating form. Then the functions $d(1,0) = d(0,1) = 1$ and $d(5,0) = d(0,5) = 1$ give the same sets $S(d,T)$ but only the first is a variation of Hodge structures. The u given naturally by the second d is the fifth power of that given by the first d, and so $u(z)$ does not act as multiplication by z on the tangent space.

SKETCH OF PROOF. (a) Let $S^+ = S(d,T)^+$. Because the Hodge filtration determines the Hodge decomposition (see (19)), the map

$$s \mapsto F_s^\bullet \colon S^+ \xrightarrow{\varphi} G_{\mathbf{d}}(V(\mathbb{C}))$$

is injective. Let G be the smallest algebraic subgroup of GL_V such that $h(\mathbb{S}) \subset G$ for all $h \in S^+$ (take G to be the intersection of the algebraic subgroups of GL_V with this property), and let $h_o \in S^+$. For all $g \in G(\mathbb{R})^+$, $gh_og^{-1} \in S^+$, and it follows from Deligne 1979, 1.1.12, that the map $g \mapsto gh_og^{-1} \colon G(\mathbb{R})^+ \to S^+$ 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 manifold (in fact, it is a real analytic manifold). Therefore, S^+ acquires the structure of a smooth manifold from

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

Let $\mathfrak{g} = \mathrm{Lie}(G)$. From the inclusion $G \hookrightarrow \mathrm{GL}(V)$ we obtain an inclusion of Lie algebras $\mathfrak{g} \hookrightarrow \mathrm{End}(V)$. This inclusion is equivariant for the adjoint action of G on $\mathfrak{g}$ and the natural action of G on $\mathrm{End}(V)$, and so h_o makes it into an inclusion of Hodge structures. Clearly, $\mathfrak{g}^{00} = \mathrm{Lie}(K_o)$ and so $\mathrm{Tgt}_{h_o}(S^+) \simeq \mathfrak{g}/\mathfrak{g}^{00}$. Consider the diagram

$$\begin{tikzcd}
\mathrm{Tgt}_{h_o}(S^+) \simeq \mathfrak{g}/\mathfrak{g}^{00} \arrow[r, hook] \arrow[d, "(24)"', "\simeq"] & \mathrm{End}(V)/\mathrm{End}(V)^{00} \arrow[d, "(24)"', "\simeq"] \\
\mathfrak{g}_{\mathbb{C}}/F^0 \arrow[r, hook] & \mathrm{End}(V(\mathbb{C}))/F^0 \simeq \mathrm{Tgt}_{h_o}(G_{\mathbf{d}}(V(\mathbb{C}))).
\end{tikzcd} \tag{26}$$

The map from top-left to bottom-right is $(d\varphi)_{h_o}$, which therefore maps $\mathrm{Tgt}_{h_o}(S^+)$ onto a complex subspace of $\mathrm{Tgt}_{h_o}(G_{\mathbf{d}}(V(\mathbb{C})))$. Since this is true for all $h_o \in S^+$, we see that φ identifies S^+ with an almost-complex submanifold $G_{\mathbf{d}}(V(\mathbb{C}))$. It can be shown that the almost-complex structure on S^+ is integrable, and so it 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) Let $h, h_o \in S^+$, so that $h = gh_og^{-1}$ for some $g \in G(\mathbb{R})^+$. Because V_h has weight n for all $h \in S^+$, $h(r)$ acts as r^{-n} on V for $r \in \mathbb{R}$. Therefore $gh_o(r)g^{-1} = h_o(r)$ for all $g \in G(\mathbb{R})^+$, and so $h_o(r) \in Z(G)$. It follows that $z \mapsto h_o(\sqrt{z})$ is a well-defined homomorphism $u_o \colon U_1 \to G^{\mathrm{ad}}$. Let $C = h_o(i) = u_o(-1)$. The faithful representation $G \to \mathrm{GL}(V)$ carries a C-polarization, namely, t_0, and so $\mathrm{ad}\,C$ is a Cartan involution on G (hence on G^{ad}) (1.20). Thus (G, u_o) satisfies condition (b) of Theorem 1.21. From the diagram (26), we see that

$$\mathfrak{g}_{\mathbb{C}}/\mathfrak{g}^{00} \simeq \mathrm{Tgt}_o(S^+) \subset \mathrm{Tgt}_o(\mathrm{Gr}_{\mathbf{d}}(V)) \simeq \mathrm{End}(V)/F^0\,\mathrm{End}(V).$$

If the family (h_s) satisfies Griffiths transversality, then $\mathfrak{g}_{\mathbb{C}}/\mathfrak{g}^{00} \subset F^{-1}\,\mathrm{End}(V)/F^0\,\mathrm{End}(V)$, and it follows that (G, u_o) satisfies condition (b) of Theorem 1.21. As (G, u_o) obviously satisfies (c), we can conclude that S^+ is the hermitian symmetric domain attached to (G, u_o).

(c) Let D be an irreducible hermitian symmetric domain, and let G be the connected adjoint group such that $G(\mathbb{R})^+ = \mathrm{Hol}(D)^+$ (see 1.7). Choose a faithful self-dual representation $G \to \mathrm{GL}(V)$ of G. Because V is self-dual, there is a nondegenerate bilinear form t_0 on

V fixed by G. Apply Proposition 2.1 to find a set of tensors T containing t_0 such that G is the subgroup of $\mathrm{GL}(V)$ fixing the $t \in T$. Let h_o be the composite of the homomorphisms

$$\mathbb{S} \xrightarrow{z \mapsto z/\bar{z}} U_1 \xrightarrow{u_o} G \to \mathrm{GL}(V),$$

where 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.[17] □

REMARK 2.15. The map $S^+ \to G_{\mathbf{d}}(V(\mathbb{C}))$ 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 $T \to G_{\mathbf{d}}(V(\mathbb{C}))$ factors into

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

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

ASIDE 2.16. Griffiths studied the variations of Hodge structures arising from smooth families of algebraic varieties in a series of papers in the 1960s. See Deligne 1970 and Voisin 2002, Chapters 9 and 10.

ASIDE. Let (V,h) be a polarizable rational Hodge structure, and let G be the smallest algebraic subgroup G of $\mathrm{GL}(V)$ such that $h(\mathbb{S}) \subset G_{\mathbb{R}}$. The pair (G,h) is called the ***Mumford-Tate group*** of (V,h).

(a) Let G be a connected algebraic group over $\mathbb{Q}$ and h a homomorphism $\mathbb{S} \to G_{\mathbb{R}}$. The pair (G,h) is the Mumford-Tate group of a rational Hodge structure if and only if G is generated by h and the weight homomorphism w_h is defined over $\mathbb{Q}$ and maps into the centre of G.

(b) A rational Hodge structure is polarizable if and only if its Mumford-Tate group (G,h) is such that $\mathrm{ad}(h(i))$ is a Cartan involution on $G/w_h(\mathbb{G}_m)$.

(c) It is also possible to characterize the abelian Hodge structures by their Mumford-Tate groups, but this is more complicated to state.

For these statements, see Milne 1994b, 1.6, 1.27; also Milne 2013, 6.2, 6.3.

ASIDE. Hodge structures can be used to give a geometric interpretation of the Borel and Harish-Chandra embeddings (Deligne), and also the Satake compactification of D.

ASIDE. We have realized the hermitian symmetric domains as parameter spaces for Hodge structures. Better, they can be realized as moduli varieties for Hodge structures (in the category of complex manifolds, for example). See Milne 2013. Realizing a hermitian symmetric domain as a parameter space for Hodge structures requires choosing a representation of the group, but this can be avoided by using the Tannakian point of view (in a sense, one choose all representations).

NOTES. Theorem 2.14 has been extracted from Deligne 1979, 1.1. There is a more complete exposition of the material in this section in Milne 2013, §5–§7.

[17] Given a pair $(V,(V^{p,q})_{p,q},T)$, define L to be the sub-Lie-algebra of $\mathrm{End}(V)$ fixing the $t \in T$, i.e., such that

$$\textstyle\sum_i t(v_1,\ldots,gv_i,\ldots,v_r) = 0.$$

Then L has a Hodge structure of weight 0. We say that $(V,(V^{p,q})_{p,q},T)$ is ***special*** if L is of type $(-1,1),(0,0),(1,-1)$. The family $S(d,T)^+$ containing $(V,(V^{p,q})_{p,q},T)$ is a variation of Hodge structures if and only if (H,T) is special.

3 Locally symmetric varieties

In this section, we study quotients of hermitian symmetric domains by discrete groups. In particular, we see that the quotient by a torsion-free arithmetic subgroup has a unique structure of an algebraic variety compatible with its complex structure. In this way, we obtain a very large class of interesting algebraic varieties over $\mathbb{C}$.

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 all $p \in D$, $\{g \in \Gamma \mid gp = p\}$ is finite;

(b) for all $p \in D$, 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 that 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 Hausdorff. 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. □

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

$$F = \left\{z \in \mathcal{H}_1 \;\middle|\; -\tfrac{1}{2} < \Re(z) < \tfrac{1}{2},\ |z| > 1\right\}$$

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 Morris 2015, 1.3, Exercise 6).

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.[18]

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[19] $G \hookrightarrow \mathrm{GL}_n$. 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. See Borel 1969, 8.9, 8.11, or Platonov and Rapinchuk 1994, Theorem 4.1, p. 204. □

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 $\mathbb{R}^\times$. It follows that 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).*

[18] If H and H' are subgroups of finite index in a group G, then $H \cap H'$ has finite index in H (because $H/H \cap H' \to G/H'$ is injective). It follows that if H_1 and H_3 are each commensurable with H_2, then $H_1 \cap H_2 \cap H_3$ has finite index in each of $H_1 \cap H_2$ and $H_2 \cap H_3$ (and therefore in H_1 and H_3). Hence, $H_1 \cap H_3$ has finite index in each of H_1 and H_3.

[19] Here, embedding means injective homomorphism.

(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*).*[20]

PROOF. See Borel 1969, 13.2, 8.4, or Platonov and Rapinchuk 1994, Theorem 4.13, p. 213, Theorem 4.12, p. 210. [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 to have no rational unipotent elements means that there are 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 $G(\mathbb{Q})=\{b\in B\mid \mathrm{Nm}(b)=1\}$ (reduced norm). 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 is isomorphic to $M_2(\mathbb{Q})$, then G is isomorphic to 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 a 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 is not isomorphic to $M_2(\mathbb{Q})$, then 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, $\Gamma\backslash G(\mathbb{R})$ is compact. In this way we get compact quotients $\Gamma\backslash\mathcal{H}_1$ of $\mathcal{H}_1$.

Let k be a subfield of $\mathbb{C}$. An automorphism of a k-vector space V is said to be ***neat*** if its eigenvalues in $\mathbb{C}$ generate a torsion-free subgroup of $\mathbb{C}^\times$. For example, no nontrivial automorphism of finite order is neat. 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 $\mathbb{C}$ (because all representations G can be constructed from one faithful representation). 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 chosen to be defined by congruence conditions (i.e., for some embedding $G\hookrightarrow\mathrm{GL}_n$ and integer N, $\Gamma'=\{g\in\Gamma\mid g\equiv 1 \bmod N\}$).*

PROOF. See Borel 1969, 17.4, or Witte Morris 2015, 4.8. □

Let H be a connected real Lie group. A subgroup Γ of H is ***arithmetic*** if there exists an algebraic group G over $\mathbb{Q}$, a surjective homomorphism $G(\mathbb{R})^+\to\mathrm{Hol}(D)^+$ with compact kernel, and an arithmetic subgroup Γ_0 of $G(\mathbb{Q})$ such that $\Gamma_0\cap G(\mathbb{R})^+$ maps onto Γ. We are only interested in the case that H is semisimple, in which case we can take G to be semisimple.

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.

[20] Recall that $\mathrm{Hom}(G,\mathbb{G}_m)=0$ means that there is no nonzero homomorphism $G\to\mathbb{G}_m$ *defined over* $\mathbb{Q}$. An element g of $G(k)$ is ***unipotent*** if $\rho(g)$ is unipotent in $\mathrm{GL}(V)$ for one faithful representation $\rho\colon G\hookrightarrow\mathrm{GL}(V)$, in which case $\rho(g)$ is unipotent for every representation ρ of G.

Because Γ_0 is discrete in $G(\mathbb{R})$, there exists an open $U \subset G(\mathbb{R})^+$ whose intersection with $\Gamma_0 \cdot \mathrm{Ker}(\alpha)$ is exactly $\mathrm{Ker}(\alpha)$. 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) to G and Γ_0. 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 subgroups 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 $\mathrm{SL}_2(\mathbb{R})$). However, in any connected real Lie group there are only countably many arithmetic subgroups up to conjugacy (Witte Morris 2015, 5.1.20).

The following major theorem of Margulis shows that SL_2 is exceptional in this regard: every discrete subgroup of finite covolume in a noncompact simple real Lie group H is arithmetic unless H is isogenous to $\mathrm{SO}(1,n)$ or $\mathrm{SU}(1,n)$ (see Witte Morris 2015, 5.2, 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 $A \otimes_k k^{\mathrm{a}}$ 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 $\mathrm{spm}(A)$ to be the set of maximal ideals in A endowed with the topology having as base the collection of sets

$$D(f) \stackrel{\text{def}}{=} \{\mathfrak{m} \mid f \notin \mathfrak{m}\}, \quad f \in A.$$

There is a unique sheaf of k-algebras $\mathcal{O}$ on $\mathrm{spm}(A)$ such that $\mathcal{O}(D(f)) = A_f$ for all f. Here A_f is the algebra obtained from A by inverting f. An ***affine algebraic variety*** over k is a ringed space isomorphic to one of the form

$$\mathrm{Spm}(A) = (\mathrm{spm}(A), \mathcal{O}).$$

The stalk at $\mathfrak{m}$ is the local ring $A_\mathfrak{m}$, and so $\mathrm{Spm}(A)$ 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 $\mathrm{spm}(A)$ 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 $\mathrm{spm}(A)$ 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. An ***algebraic variety*** over k is a separated prevariety over k.

Alternatively, the algebraic varieties over k are precisely the ringed spaces obtained from geometrically-reduced separated schemes of finite type over k by deleting the nonclosed points.[21]

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.

Let V be an algebraic variety over k, and let R be a k-algebra. We let $V(R)$ denote the set of points of V with coordinates in R. When A is affine, $V(R) = \mathrm{Hom}_{k\text{-algebra}}(A, R)$. In general, $V(R) = \mathrm{Hom}_k(\mathrm{Spm}(R), V)$.

Let V be an algebraic variety over k. The ***tangent space*** $\mathrm{Tgt}_p(V)$ to V at a point $p \in V(k)$ is the k-vector space of k-derivations $\mathcal{O}_{V,p} \to k$. A point $p \in V(k)$ is said to be ***nonsingular*** if the dimension of $\mathrm{Tgt}_p(V)$ is equal to the dimension of the connected component of V containing p. The variety V is ***nonsingular*** if every $p \in V(k)$ is nonsingular and this remains true when we extend scalars to k^{a}. A map of nonsingular varieties is ***étale*** if it induces isomorphisms on the tangent spaces and this remains true when we extend scalars to k^{a}.

For the variety approach to algebraic geometry, see my notes *Algebraic Geometry*, and for the scheme approach, see Hartshorne 1977.

Algebraic varieties versus complex manifolds

For a nonsingular algebraic variety V over $\mathbb{C}$, the set $V(\mathbb{C})$ has a natural structure as a complex manifold. More precisely, there is the following statement.

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

(a) *$V = V^{\mathrm{an}}$ as sets, every Zariski-open subset is open for the complex topology, and every regular function is holomorphic;*[22]

(b) *if $V = \mathbb{A}^n$, then $V^{\mathrm{an}} = \mathbb{C}^n$ with its natural structure as a complex manifold;*

(c) *if $\varphi \colon V \to W$ is étale, then $\varphi^{\mathrm{an}} \colon V^{\mathrm{an}} \to W^{\mathrm{an}}$ is a local isomorphism.*

PROOF. A regular map $\varphi \colon V \to W$ is étale if the map $d\varphi_p \colon \mathrm{Tgt}_p V \to \mathrm{Tgt}_p W$ 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, 5.53), this shows that there exists at most one functor satisfying (a,b,c), and suggests how to define it. □

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 $V \rightsquigarrow V^{\mathrm{an}}$ is faithful, but it is neither full nor essentially surjective on objects.

[21] Nonclosed points are not needed when working with algebraic varieties over a field, and so algebraic geometers often ignore them. For example, it makes the statement of Proposition 3.8 simpler when we do.

[22] These conditions require that the identity map $V \to V$ be a morphism of $\mathbb{C}$-ringed spaces $(V^{\mathrm{an}}, \mathcal{O}_{V^{\mathrm{an}}}) \to (V, \mathcal{O}_V)$. This morphism is universal.

REMARK 3.9. The functor $V \rightsquigarrow V^{\mathrm{an}}$ can be extended to all algebraic varieties once one has the notion of a "complex manifold with singularities". This is called a ***complex analytic 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).

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 (Hironaka 1964).[23] 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

$$\{\text{projective nonsingular curves over } \mathbb{C}\} \rightsquigarrow \{\text{compact Riemann surfaces}\}$$

is an equivalence of categories (see MF, pp. 94–95, 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 3.10(a) 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 every 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 all lattices 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.[24] For quotients of $\mathbb{C}^g$ by a lattice Λ, condition 3.10(b) is sufficient for algebraicity (Mumford 1970, p. 35).

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:

[23] For a reasonably elementary exposition of resolution in characteristic zero, see Kollár 2007.

[24] A complex torus $\mathbb{C}^g/\Lambda$ is algebraic if and only if it admits a Riemann form (see 6.7 below). When $g > 1$, those that admit a Riemann form are a proper closed subset of the moduli space. If Λ is the lattice in $\mathbb{C}^2$ generated by $(1,0), (i,0), (0,1), (\sqrt{2}, i)$, then $\mathbb{C}^2/\Lambda$ does not admit a Riemann form (Shafarevich 1994, VIII 1.4).

THEOREM 3.11 (CHOW 1949). *Every projective complex analytic space has a unique structure of a projective algebraic variety, and every holomorphic map of projective complex analytic spaces is regular for these structures. The algebraic variety attached to a complex manifold is nonsingular.*

PROOF. See Shafarevich 1994, VIII 3.1, for a proof in the nonsingular case. □

In other words, the functor $V \rightsquigarrow V^{\mathrm{an}}$ is an equivalence from the category of projective algebraic varieties to the category of projective complex analytic spaces under which nonsingular algebraic varieties correspond to complex manifolds.

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 D 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 of 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 $\Gamma\backslash D^*$.

Here is some of the history. 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.[25]

[25] For a discussion of later work, see Casselman 1997.

REMARK 3.13. (a) The theorem also holds when Γ has torsion. Then $\Gamma \backslash D$ is a normal complex analytic space (rather than a manifold) and it has the structure of a normal algebraic variety (rather than a nonsingular algebraic variety).

(b) 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.

(c) The variety $D(\Gamma)^*$ is equal to $\mathrm{Proj}(\bigoplus_{n\geq 0} A_n)$, where A_n is the vector space of automorphic forms for the nth 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

$$D(\Gamma)^* = \mathrm{Proj}(\textstyle\bigoplus_{n\geq 0} H^0(D(\Gamma), \omega^n)),$$

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).

(c) 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 $D(\Gamma)$ be the quotient of a hermitian symmetric domain D by a torsion-free arithmetic subgroup Γ of $\mathrm{Hol}(D)^+$, and let V be a nonsingular quasi-projective variety over $\mathbb{C}$. Then every holomorphic map $f: V^{\mathrm{an}} \to D(\Gamma)^{\mathrm{an}}$ is regular.*

Let $D(\Gamma)^*$ be as in 3.12. The key step in Borel's proof is the following result:

LEMMA 3.15. *Let $\mathcal{D}_1^\times$ denote the punctured disk $\{z \mid 0 < |z| < 1\}$. Then every holomorphic map*[26] *$\mathcal{D}_1^{\times r} \times \mathcal{D}_1^s \to D(\Gamma)$ 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 holomorphic 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 $\mathbb{P}^1(\mathbb{C}) \smallsetminus \{3 \text{ points}\}$. 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 $\mathcal{D}_1^{\times r} \times \mathcal{D}_1^s$.

[26] 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.

Using the lemma, we can prove the theorem. According to Hironaka's theorem on the resolution of singularities (see p. 37), 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, and is therefore an isomorphism (AG 8.60). □

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

$$\begin{tikzcd} D(\Gamma) \arrow[r] \arrow[dr] & D(\Gamma)^\dagger \arrow[d, dashed] \\ & D(\Gamma)^* \end{tikzcd}$$

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) The statement of Theorem 3.14 also holds for singular V. Let $f\colon V^{\mathrm{an}} \to D(\Gamma)^{\mathrm{an}}$ be holomorphic. Then Theorem 3.14 shows that f becomes regular when restricted to the complement of the singular locus in V, which is open and dense, and this implies that f is regular on V.

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

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 nontrivial normal subgroup of compact type.

Let G be a semisimple group G over $\mathbb{Q}$. There is an isogeny $G_1 \times \cdots \times G_r \to G$ with each G_i simple (Milne 2017, 21.51). The group G is of compact type if every $G_i(\mathbb{R})$ is compact and of noncompact type if no $G_i(\mathbb{R})$ is compact. 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. See Borel 1969, 15.12, or Witte Morris 2015, 4.5, 4.6 — the hypothesis is obviously necessary. □

COROLLARY 3.20. *Let G be as in the theorem, and let Z be its centre (as an algebraic group over $\mathbb{Q}$). The centralizer in $G(\mathbb{R})$ of any arithmetic subgroup Γ of $G(\mathbb{Q})$ is $Z(\mathbb{R})$.*

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 of $\mathrm{Hol}(D)^+$. Then $D(\Gamma)$ has only finitely many automorphisms (as a complex manifold).*

PROOF. As Γ is torsion-free, D is the universal covering space of $\Gamma\backslash D$ and Γ is the group of covering transformations (see p. 32). An automorphism $\alpha: \Gamma\backslash D \to \Gamma\backslash D$ lifts to an automorphism $\tilde{\alpha}: D \to D$. For all $\gamma \in \Gamma$, the map $\tilde{\alpha}\gamma\tilde{\alpha}^{-1}$ is a covering transformation, and so lies in Γ. Conversely, an automorphism of D normalizing Γ defines an automorphism of $\Gamma\backslash D$. Thus,

$$\mathrm{Aut}(\Gamma\backslash D) = N(\Gamma)/C(\Gamma),$$

where $N(\Gamma)$ (resp. $C(\Gamma)$) is the normalizer (resp. centralizer) of Γ in $\mathrm{Aut}(D)$.

By assumption, there exists a semisimple algebraic group G over $\mathbb{Q}$, a surjective homomorphism $G(\mathbb{R})^+ \to \mathrm{Hol}(D)^+$ with compact kernel, and an arithmetic subgroup Γ_0 of $G(\mathbb{Q})$ such that $\Gamma_0 \cap G(\mathbb{R})^+$ maps onto Γ with finite kernel. We may discard any compact isogeny factors of G, and so suppose that G is of noncompact type (apply (a)⇔(c) of 4.7 below if this isn't obvious). Let N^+ be the identity component of $N(\Gamma)$. Because Γ is discrete, N^+ acts trivially on it, and so N^+ is contained in the (finite) centre of $G(\mathbb{R})$. Therefore $N(\Gamma)$ is discrete. Because $\Gamma\backslash\mathrm{Aut}(D)$ has finite volume (3.3a), this implies that Γ has finite index in $N(\Gamma)$ (cf. Witte Morris 2015, 4.5.5).

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. Mostly in this section, we have required Γ to be torsion-free. In particular, we disallowed $\Gamma(1)\backslash\mathcal{H}_1$. For an arithmetic subgroup Γ with torsion, the algebraic variety $\Gamma\backslash D$ may be singular and Borel's theorem 3.14 fails. But see the asides p. 75 and p. 94.

NOTES. Witte Morris 2015 is a friendly introduction to the theory of discrete subgroups of Lie groups. See also Borel 1969, Raghunathan 1972, and Platonov and Rapinchuk 1994. There is a large literature on the various compactifications of locally symmetric varieties. For overviews, see Satake 2001 and Goresky 2005, 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. See also Borel and Ji 2006.

4 Connected Shimura varieties

In the last chapter, we saw (3.12) that the quotient $\Gamma\backslash D$ of a hermitian symmetric domain D by a torsion-free arithmetic subgroup Γ of $\mathrm{Hol}(D)^{+}$ is an algebraic variety $D(\Gamma)$. In this chapter, we study those varieties $D(\Gamma)$ for which Γ is defined by congruence conditions.

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

$$\Gamma(N) = \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \;\middle|\; a,d \equiv 1, \quad b,c \equiv 0 \quad \bmod N\right\}.$$

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 of 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[27] SL_2 and all nonsimply connected groups have many noncongruence arithmetic subgroups (see A.3). In contrast to arithmetic subgroups, the image of a congruence subgroup under an isogeny of algebraic groups need not be a congruence subgroup. For more on congruence subgroups, see A.3.

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

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

where ℓ runs over the finite primes of $\mathbb{Q}$ (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 $\prod_\ell \mathbb{Z}_\ell$ is open and has the product topology. Note that $\hat{\mathbb{Z}} \overset{\text{def}}{=} \varprojlim_n \mathbb{Z}/n\mathbb{Z}$ equals $\prod_\ell \mathbb{Z}_\ell$, and that $\mathbb{A}_f = \hat{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathbb{Q}$.

Let $V = \mathrm{Spm}(A)$ be an affine variety over $\mathbb{Q}$. The set of points of V with coordinates in a $\mathbb{Q}$-algebra R is $V(R) = \mathrm{Hom}_{\mathbb{Q}}(A, R)$. 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

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

[27] That $\mathrm{SL}_2(\mathbb{Z})$ has noncongruence arithmetic subgroups was first noted in Klein 1880. For a proof that $\mathrm{SL}_2(\mathbb{Z})$ has infinitely many subgroups of finite index that are not congruence subgroups see Sury 2003, 3-4.1. The proof proceeds by showing that the groups occurring as quotients of $\mathrm{SL}_2(\mathbb{Z})$ by congruence subgroups (especially by principal congruence subgroups) are of a rather special type, and then exploits the known structure of $\mathrm{SL}_2(\mathbb{Z})$ as an abstract group to construct many finite quotients not of this type. In fact, the proportion of subgroups of index dividing m in $\mathrm{SL}_2(\mathbb{Z})$ that are congruence subgroups tends to zero with increasing m (Stothers 1984).

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

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

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 all 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, the restricted topological product

$$V(\mathbb{A}_f)\overset{\text{def}}{=}\prod\nolimits_\ell(V(\mathbb{Q}_\ell),V(\mathbb{Z}_\ell))$$

is independent of the choice of generators for A.[28]

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

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

similarly. This is a topological group. For example,

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

PROPOSITION 4.1. *Let K be a compact open subgroup of $G(\mathbb{A}_f)$. Then $K\cap G(\mathbb{Q})$ is a congruence subgroup of $G(\mathbb{Q})$, and every congruence subgroup is of this form.*

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

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

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

[28] 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 $\mathrm{Spec}\,\mathbb{Z}$. A different closed immersion β gives a different flat model V_β, but for some d, the isomorphism $(V_\alpha)_{\mathbb{Q}}\simeq V\simeq(V_\beta)_{\mathbb{Q}}$ on generic fibres extends to an isomorphism $V_\alpha\to V_\beta$ over $\mathrm{Spec}\,\mathbb{Z}[\frac{1}{d}]$. 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.

More generally, an arbitrary variety V over $\mathbb{Q}$ has flat models over $\mathrm{Spec}\,\mathbb{Z}$, any two of which become isomorphic over a nonempty open subset of $\mathrm{Spec}\,\mathbb{Z}$. This allows one to define $V(\mathbb{A}_f)$.

ASIDE. A basic compact open subgroup K of $G(\mathbb{A}_f)$ is defined by imposing a congruence condition at each of a finite set of primes. Then $\Gamma = G(\mathbb{Q}) \cap K$ is obtained from $G(\mathbb{Z})$ by imposing the same congruence conditions. You should think of Γ as being the congruence subgroup defined by the "congruence condition" K.

REMARK 4.2. There is a topology on $G(\mathbb{Q})$ for which the congruence subgroups form a fundamental system of neighbourhoods of 1. 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 all $u \in D$, only the characters $z, 1, z^{-1}$ occur in the representation of U_1 on $\mathrm{Lie}(G^{\mathrm{ad}})_{\mathbb{C}}$ defined by $\mathrm{Ad} \circ u$;

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

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 connected 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 Borel density theorem (3.19) and also strong approximation theorem when G is simply connected (see 4.16 below).

[The next statement should be moved earlier, but this would upset the numbering.]

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)$\Leftrightarrow$(b). If $u(-1)=1$, then u factors through $U_1 \xrightarrow{2} U_1$, and so $z^{\pm 1}$ cannot 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)$\Leftrightarrow$(c). We have

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

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 $G^{\mathrm{ad}}(\mathbb{R})^+ \to \mathrm{Hol}(D)^+$ 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 $G^{\mathrm{ad}}(\mathbb{R})^+ \to \mathrm{Hol}(D)^+$ 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 $H_{\mathrm{nc}}(\mathbb{R})^+ \simeq \mathrm{Hol}(D)^+$, 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

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

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

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 $G^{\mathrm{ad}}_{\mathbb{R}}$. 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_c is connected and $H_{\mathrm{c}}(\mathbb{R})$ is compact, the latter is connected,[29] and so the last part of the proposition follows from (1.7). $\square$

[29] Every compact subgroup of $\mathrm{GL}_n(\mathbb{R})$ is algebraic (Milne 2017, 9.30). Let G be a connected algebraic group over $\mathbb{R}$ such that $G(\mathbb{R})$ is compact. By the first remark, there is an algebraic subgroup H of G such that $H(\mathbb{R}) = G(\mathbb{R})^+$. As $G(\mathbb{R})^+$ is Zariski dense in G, we have $H = G$ and so $G(\mathbb{R})$ is connected.

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 the map $G^{\mathrm{ad}}(\mathbb{R})^+ \to \mathrm{Hol}(D)^+$ has compact kernel, the image $\bar{\Gamma}$ of any arithmetic subgroup Γ of $G^{\mathrm{ad}}(\mathbb{Q})^+$ in $\mathrm{Hol}(D)^+$ is arithmetic (by definition p. 34), and the kernel of $\Gamma \to \bar{\Gamma}$ is finite. If $\bar{\Gamma}$ is torsion-free, then the Baily-Borel and Borel theorems (3.12, 3.14) apply to $D(\Gamma) \stackrel{\text{def}}{=} \Gamma\backslash D = \bar{\Gamma}\backslash D$. In particular, $D(\Gamma)$ has a unique structure of an algebraic variety and, for every $\Gamma' \subset \Gamma$, the natural map

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

is regular.

DEFINITION 4.10. Let (G,D) be a connected Shimura datum. A ***connected Shimura variety*** relative to (G,D) is an algebraic variety of the form $D(\Gamma)$ with Γ an arithmetic subgroup of $G^{\mathrm{ad}}(\mathbb{Q})^+$ containing the image of congruence subgroup of $G(\mathbb{Q})^+$ and such that $\bar{\Gamma}$ is torsion-free.[30] The inverse system of such algebraic varieties, denoted $\mathrm{Sh}^{\circ}(G,D)$, is called the ***connected Shimura variety*** attached to (G,D).

REMARK 4.11. (a) Let (G,D) be a connected Shimura datum, and let τ denote the topology on $G^{\mathrm{ad}}(\mathbb{Q})$ for which the images of the congruence subgroups of $G(\mathbb{Q})$ form a fundamental system of neighbourhoods of 1. The connected Shimura varieties relative to (G,D) are the quotients $\Gamma\backslash D$ where Γ is an arithmetic subgroup of $G(\mathbb{Q})^+$ open for the topology τ and such that $\bar{\Gamma}$ is torsion-free. 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 regular (3.14). Conjugation by g on $G^{\mathrm{ad}}(\mathbb{Q})^+$ is a homeomorphism for the τ topology, and so the group $G^{\mathrm{ad}}(\mathbb{Q})^+$ acts on the inverse system $\mathrm{Sh}^{\circ}(G,D)$. This action extends by continuity to an action of the completion $G^{\mathrm{ad}}(\mathbb{Q})^{+\wedge}$ of $G^{\mathrm{ad}}(\mathbb{Q})^+$ for the τ topology (Deligne 1979, 2.1.8).

(b) The varieties $\Gamma\backslash D$ with Γ a congruence subgroup of $G(\mathbb{Q})^+$ are cofinal in the inverse system $\mathrm{Sh}^{\circ}(G,D)$.

PROPOSITION 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 $G(\mathbb{Q})^+$.*

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

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

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

[30] We can allow $\bar{\Gamma}$ to have torsion, but then we have to remember that the variety may be singular and Borel's theorem (3.14) may fail.

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. $\square$

REMARK 4.13. The homomorphism $\pi\colon G(\mathbb{Q})^+ \to G^{\mathrm{ad}}(\mathbb{Q})^+$ is usually far from surjective. For example, the image of $\mathrm{SL}_2(\mathbb{Q}) = \mathrm{SL}_2(\mathbb{Q})^+ \to \mathrm{PGL}_2(\mathbb{Q})^+$ consists of the elements represented by a matrix with determinant in $\mathbb{Q}^{\times 2}$. 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) Let $G = \mathrm{SL}_2$ and $D = \mathcal{H}_1$. 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) Let $G = \mathrm{PGL}_2$ and $D = \mathcal{H}_1$. This is 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). For a discussion of the congruence subgroups of $\mathrm{PGL}_2(\mathbb{Q})$, see Hsu 1996.

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

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

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, it is the congruence subgroups that arise naturally when we work adèlically.

The strong approximation theorem

Recall that a semisimple group G is said to be simply connected if every 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. See Platonov and Rapinchuk 1994, Theorem 7.12, p. 427. □

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

(a) $\mathbb{G}_m$ (not semisimple): the group $\mathbb{Q}^\times$ is not dense in $\mathbb{A}_f^\times$.[31]

(b) PGL_2 (not simply connected): 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: if $G(\mathbb{Q})$ were dense in $G(\mathbb{A}_f)$, then $G(\mathbb{Z}) = G(\mathbb{Q}) \cap G(\hat{\mathbb{Z}})$ would be dense in $G(\hat{\mathbb{Z}})$, but $G(\mathbb{Z})$ is discrete in $G(\mathbb{R})$ (see 3.3) and hence is finite.

An adèlic description of $D(\Gamma)$

Let G be a simply connected semisimple group over $\mathbb{Q}$. Then $G(\mathbb{R})$ is connected (5.2 below), and so $G(\mathbb{Q}) \subset G(\mathbb{R})^+ = G(\mathbb{R})$. In the next proposition, $G(\mathbb{Q})$ acts on D through the action of $G(\mathbb{R})^+$.

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*

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

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

$$\begin{tikzcd} D \arrow[r, "x \mapsto (x,[1])"] \arrow[d] & D \times (G(\mathbb{A}_f)/K) \arrow[d] \\ \Gamma\backslash D \arrow[r, "{[x] \mapsto [x,1]}"] & G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)/K. \end{tikzcd}$$

[31] Let $(a_\ell)_\ell$ be an arbitrary element of $\prod \mathbb{Z}_\ell^\times \subset \mathbb{A}_f^\times$ and let S be a finite set of prime numbers. If $\mathbb{Q}^\times$ is dense, then there exists an $a \in \mathbb{Q}^\times$ that is close to a_ℓ for $\ell \in S$ and an ℓ-adic unit for $\ell \notin S$. But such an a is an ℓ-adic unit for all ℓ, and so equals ± 1. This yields a contradiction. A similar argument shows that $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ is not dense in $\mathbb{A}_f^\times/\mathbb{A}_f^{\times 2}$.

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, as follows.

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.*

(a) *The canonical map $h\colon X/\bigcap G_i \to \varprojlim X/G_i$ is continuous.*
(b) *The map h is injective if the stabilizer in G_i of x is compact for every $x \in X$ and $i \in I$.*[32]
(c) *The map h is surjective if the orbit xG_i is compact for every $x \in X$ and $i \in I$.*

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.

(a) Let $I = \bigcap G_i$. Then I acts continuously on X, and the mapping $X/I \to X/G_i$ is continuous for every i. The inverse limit of these continuous mappings is continuous.

(b) Let $x, x' \in X$. For each i, let

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

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

(c) Let $(x_iG_i)_{i\in I} \in \varprojlim X/G_i$. Then $\varprojlim x_iG_i$ is nonempty because each orbit is compact. If $x \in \varprojlim x_iG_i$, then $x \cdot \bigcap G_i$ maps to $(x_iG_i)_{i\in I}$. □

[32]This following example shows that it is not enough to assume that every G_i is compact. Let $G = \mathbb{Z}_p$ act on $X = \mathbb{Z}_p/\mathbb{Z}$ by translation, and let $G_i = p^i\mathbb{Z}_p$. Then $X/\bigcap G_i = \mathbb{Z}_p/\mathbb{Z}$ but $X/G_i = \{0\}$ for all i.

Note that the conclusions of the lemma hold if every subgroup G_i is compact and every orbit xG_i is Hausdorff (because the hypotheses do).

We now prove Proposition 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 Hausdorff 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[33] $g \in G(\mathbb{Q}) \smallsetminus \{1\}$, 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) Let G be simply connected. Why replace the single coset space on the left of (28) with the more complicated double coset space on the right? One reason is the description of the limit

$$\varprojlim \Gamma \backslash D \simeq G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)$$

in 4.19. Another 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 when 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 inverse system $\mathrm{Sh}^\circ(G,D) = (\Gamma\backslash D)_\Gamma$ exists as a scheme[34]. It is even locally noetherian and regular, but not, of course, of finite type over $\mathbb{C}$, and it is possible to recover the inverse system from the inverse limit. Thus, it is legitimate to replace the inverse system with its limit (as Deligne does, 1979, 2.1.8). Compare 5.30 below. The map $\mathrm{Sh}^\circ \to \Gamma\backslash D$ can be regarded as an algebraic approximation to the universal covering map $D \to \Gamma/D$. Indeed, just as $G(\mathbb{R})^+$ acts on D (but not on $\Gamma\backslash D$), the τ-completion of $G(\mathbb{Q})^+$ acts on Sh° (but not on $\Gamma\backslash D$). Moreover, Sh° does behave as though it is simply connected. For example, every finite étale map of connected Shimura varieties compatible with an isomorphism of the finite-adèlic groups is an isomorphism (Milne 1983, 2.1).

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

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

[33]Let $g \in G(\mathbb{Q})$, and suppose that $g(V \times aK) \cap (V \times bK) \neq \emptyset$. Then

$$gaK = bK = aK$$

and so $g \in G(\mathbb{Q}) \cap aKa^{-1} = \Gamma$. As $gV \cap V \neq \emptyset$, this implies that $g = 1$.

[34]Let $(A_i)_{i \in I}$ be a direct system of commutative rings indexed by a directed set I, and let $A = \varinjlim A_i$. Then, for any scheme X,

$$\mathrm{Hom}(X, \operatorname{Spec} A) \simeq \mathrm{Hom}(A, \Gamma(X, \mathcal{O}_X)) \simeq \varprojlim \mathrm{Hom}(A_i, \Gamma(X, \mathcal{O}_X)) \simeq \varprojlim \mathrm{Hom}(X, \operatorname{Spec} A_i).$$

Here the middle isomorphism comes from the definition of a direct limit and the end isomorphisms from the adjointness of Spec and Γ (see Hartshorne 1977, II, Exercise 2.4). This shows that $\operatorname{Spec} A$ is the inverse limit of the inverse system $(\operatorname{Spec} A_i)_{i \in I}$ in the category of schemes. More generally, inverse limits of schemes in which the transition morphisms are affine exist, and can be constructed in the obvious way.

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 all $h \in X^{+}$, 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 $\mathrm{Ad} \circ h$;

SV2: for all $h \in X^{+}$, $\mathrm{ad}(h(i))$ is a Cartan involution on $G^{\mathrm{ad}}_{\mathbb{R}}$.

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. There Deligne often writes (2.1.1), (2.1.2), (2.1.3) when he means (2.1.1.1), (2.1.1.2), (2.1.1.3). He denotes a connected Shimura datum by a triple $(G^{\mathrm{ad}}, G, X^{+})$, but the first term is superfluous, and so we omit it.

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 on a realization of G as the derived group of a reductive group.[35] 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.[36]

How can we do this? Consider the complex line $Y+i=0$ in $\mathbb{C}^2$. 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 $(Y-i)(Y+i)=0$.[37] 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.[38] Shimura varieties give a systematic way of doing this for connected Shimura varieties.

Notation 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 $Z = Z(G)$. 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.

---

[35]Because realizing the variety as a moduli variety requires having a reductive group. Consider for example the simplest Shimura variety, with $G=\mathrm{GL}_2$ and $X=\mathcal{H}_1^{\pm}$. From a representation $\rho\colon \mathrm{GL}_2 \to V$ of GL_2 we obtain a map $h \mapsto \rho_{\mathbb{R}} \circ h$ from X to the set of rational polarizable Hodge structures. When we take ρ to be the standard representation of GL_2, this becomes a map from X to elliptic curves. For the connected Shimura datum (G', X^+) with $G' = \mathrm{SL}_2$, the homomorphisms h have target $G'^{\mathrm{ad}}_{\mathbb{R}}$, and so only representations of G'^{ad}, not G', give rational polarizable Hodge structures. This is not what we want.

[36]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 1981b. See also Appendix C.

[37]This should not be confused with Weil restriction of scalars. For example, the variety over $\mathbb{R}$ obtained from the line $Y+i=0$ by restriction of scalars is two-dimensional and isomorphic to $\mathbb{A}^2$.

[38]Let V be a connected nonsingular variety over a field k of characteristic zero. Then V is geometrically connected (i.e., $V \otimes_k k^{\mathrm{a}}$ is connected) if and only if k is algebraically closed in $\Gamma(V, \mathcal{O}_V)$. Suppose that V is geometrically connected, and let k_0 be a subfield of k such that $[k:k_0] < \infty$. Then V can also be regarded as a k_0-variety (same V, same $\mathcal{O}_V$ but regarded as a sheaf of k_0-algebras), and

$$V \otimes_{k_0} k^{\mathrm{a}} \simeq \coprod\nolimits_{\sigma} V \otimes_{k,\sigma} k^{\mathrm{a}}$$

where σ runs through the k_0-embeddings of k into k^{a}.

For a reductive group G, there is a diagram

$$\begin{tikzcd} & G^{\mathrm{der}} \arrow[d] \arrow[dr] & \\ Z \arrow[r] \arrow[dr] & G \arrow[r, "\mathrm{ad}"] \arrow[d, "\nu"] & G^{\mathrm{ad}} \\ & T & \end{tikzcd}$$

in which the column and row are short exact sequences, the diagonal maps are isogenies with kernel the centre $Z \cap G^{\mathrm{der}}$ of G^{der}, and T (a torus) is the largest commutative quotient of G. This gives rise to a short exact sequence

$$1 \to Z \cap G^{\mathrm{der}} \to Z \times G^{\mathrm{der}} \to G \to 1. \tag{30}$$

For example, when $G = \mathrm{GL}_n$, the diagrams become

$$\begin{tikzcd} & \mathrm{SL}_n \arrow[d] \arrow[dr] & \\ \mathbb{G}_m \arrow[r] \arrow[dr, "x \mapsto x^n"'] & \mathrm{GL}_n \arrow[r, "\mathrm{ad}"] \arrow[d, "\det"] & \mathrm{PGL}_n \\ & \mathbb{G}_m & \end{tikzcd}$$

and

$$1 \to \mu_n \to \mathbb{G}_m \times \mathrm{SL}_n \to \mathrm{GL}_n \to 1.$$

There is an action of G^{ad} on G for which $\mathrm{ad}(g)$, $g \in G(\mathbb{Q})$, acts as $x \mapsto gxg^{-1}$.

The cohomology set $H^1(\mathbb{Q}, G)$ is defined to be the set of continuous crossed homomorphisms $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q}) \to G(\mathbb{Q}^{\mathrm{a}})$ modulo the relation that identifies two crossed homomorphisms differing by a principal crossed homomorphism (more precisely, $f' \sim f$ if there exists an $a \in G(\mathbb{Q}^{\mathrm{a}})$ such that $f'(\sigma) = a^{-1} \cdot f(\sigma) \cdot \sigma(a)$ for all $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$). It is a set with a distinguished element e, represented by any principal crossed homomorphism. An element of $T(\mathbb{Q})$ lifts to an element of $G(\mathbb{Q})$ if and only if it maps to the distinguished class in $H^1(\mathbb{Q}, G^{\mathrm{der}})$.

The real points of algebraic groups

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

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 $H(\mathbb{R})^+$. $\square$

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). *Let G be a semisimple algebraic group over $\mathbb{R}$. If G is simply connected group G, then $G(\mathbb{R})$ is connected.*

PROOF. See Platonov and Rapinchuk 1994, Theorem 7.6, p. 407. □

COROLLARY 5.3. *For a reductive group G over $\mathbb{R}$, $G(\mathbb{R})$ has only finitely many connected components (for the real topology).*[39]

PROOF. It follows from Proposition 5.1 that an exact sequence of real algebraic groups

$$1 \to N \to G' \to G \to 1 \tag{31}$$

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 simply connected covering group of G^{der}. As G is an almost direct product of $Z = Z(G)$ and G^{der} (30), there is an exact sequence (31) 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 $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ is finite and 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). *Let G be a connected algebraic group over $\mathbb{Q}$. Then $G(\mathbb{Q})$ is dense in $G(\mathbb{R})$.*

PROOF. As Deligne (1971b, 0.4) writes “on se ramène aisément au cas des tores”. This is explained in Addendum A. □

More generally, $G(\mathbb{Q})$ is dense in $G(\mathbb{R})$ if every connected component of G contains $\mathbb{Q}$-point.

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 following conditions:

[39]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, p. 119.

SV1: for all $h \in X$, the Hodge structure on $\mathrm{Lie}(G_{\mathbb{R}})$ defined by $\mathrm{Ad} \circ h$ is of type

$$\{(-1,1),(0,0),(1,-1)\};$$

SV2: for all $h \in X$, $\mathrm{ad}(h(i))$ is a Cartan involution of $G_{\mathbb{R}}^{\mathrm{ad}}$;

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

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). Note also that SV1 says that, for $h \in X$, only the characters $z/\bar{z}$, 1, $\bar{z}/z$ occur in the representation of $\mathbb{S}$ on $\mathrm{Lie}(G_{\mathbb{C}})$ defined by $\mathrm{Ad} \circ h$. As $\mathrm{Lie}(G_{\mathbb{C}}) = \mathrm{Lie}(Z_{\mathbb{C}}) \oplus \mathrm{Lie}(G_{\mathbb{C}}^{\mathrm{ad}})$ and $\mathrm{Ad}((h(z))$ automatically acts trivially on $\mathrm{Lie}(Z_{\mathbb{C}})$, this is the same as the earlier condition SV1.

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 $z/\bar{z}$ (rather than $\bar{z}/z$).

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

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

(b) *Let X^+ be a connected component of X, and let $\bar{X}^+$ be its image in $\bar{X}$. If (G,X) satisfies the axioms SV1–3 then $(G^{\mathrm{der}}, \bar{X}^+)$ 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 $h \mapsto \bar{h} \colon X \to \bar{X}$ is injective. For the second part of the statement, use that $G^{\mathrm{ad}}(\mathbb{R})^+$ acts transitively on each connected component of $\bar{X}$ (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 $\pi_0(X) \subset \pi_0(\bar{X})$. The stabilizer in $G^{\mathrm{ad}}(\mathbb{R})$ of $[\bar{X}^+]$ 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 $G(\mathbb{R})$. □

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. □

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 $z/\bar{z}$, 1, $\bar{z}/z$. 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***. Let $\rho\colon G_{\mathbb{R}}\to\mathrm{GL}(V)$ be a representation; then $\rho\circ w_X$ defines a decomposition $V=\bigoplus V_n$ of V 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 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. As every representation of G can be constructed out of one faithful representation (Milne 2017, 4.14), the family of Hodge structures defined by any representation is then holomorphic for this complex structure. The condition SV1 implies that $(V,\rho\circ h)_h$ is a variation of Hodge structures (see the proof of the converse in Theorem 1.21). Now (b) of Theorem 2.14 implies that the connected components of X are hermitian symmetric domains. □

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 and let $\bar{X}$ be as in 5.7. The maps

$$\pi_0(X)\to\pi_0(\bar{X})$$
$$G(\mathbb{R})/G(\mathbb{R})_+\to G^{\mathrm{ad}}(\mathbb{R})/G^{\mathrm{ad}}(\mathbb{R})^+$$

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 every 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 every 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 (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 a compact open subgroup K of $G(\mathbb{A}_f)$, consider the double coset space[40]

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

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

$$G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K \simeq \bigsqcup\nolimits_{g\in\mathcal{C}} \Gamma_g\backslash X^+,$$

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. For $g \in \mathcal{C}$, consider the map

$$[x] \mapsto [x,g]\colon \Gamma_g\backslash X^+ \to G(\mathbb{Q})_+\backslash X^+ \times G(\mathbb{A}_f)/K. \tag{32}$$

I claim that, for each g, the map (32) 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 for differing g. Therefore, the first statement of the lemma follows from (5.11), and the second statement can be proved in the same way as the similar statement in (4.18).

To see that the map (32) is injective, note that if $[x,g] = [x',g]$, then $x' = qx$ and $g = qgk$ for some $q \in G(\mathbb{Q})_+$ and $k \in K$. From the second equation, we find that $q \in \Gamma_g$, and so $[x] = [x']$.

For the second part of the claim, let $(x,a) \in G(\mathbb{A}_f)$. Then $a = qgk$ for some $q \in G(\mathbb{Q})_+$, $g \in \mathcal{C}$, $k \in K$. Now $[x,a] = [q^{-1}x, g]$, which lies in the image of $\Gamma_g\backslash X^+$. Suppose that $[x,g] = [x',g']$ with $g, g' \in \mathcal{C}$. Then $x' = qx$ and $g' = qgk$ for some $q \in G(\mathbb{Q})_+$ and $k \in K$. The second equation implies that $g' = g$. □

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

[40] Let $\mathbb{A} = \mathbb{R} \times \mathbb{A}_f$ be the full ring of adèles, and let K_∞ denote the centralizer of h in $G(\mathbb{R})$. Then

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

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. Let (G,X) be a Shimura datum. A ***Shimura variety*** relative to (G,X) is a variety of the form $\mathrm{Sh}_K(G,X)$ for some (small) compact open subgroup K of $G(\mathbb{A}_f)$. The ***Shimura variety*** $\mathrm{Sh}(G,X)$ attached to[41] a 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 $G(\mathbb{A}_f)$.

Thus, a Shimura variety relative to (G,X) is a finite disjoint union of arithmetic locally symmetric varieties, and the Shimura variety attached to (G,X) is an inverse system of such varieties equipped with an action 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 Theorem 3.14, and the second is proved in Theorem 1.15 of Deligne 1971b. □

ASIDE. The second part of the theorem requires explanation. It says that $\mathrm{Sh}(G,X) \to \mathrm{Sh}(G',X')$ is a closed immersion as a morphism of inverse systems of algebraic varieties (or in the inverse limit; see 5.30). Specifically, for every sufficiently small compact open subgroup K' of $G'(\mathbb{A}_f)$, there is a compact open subgroup K of $G(\mathbb{A}_f)$ such that the map

$$\mathrm{Sh}_K(G,X) \to \mathrm{Sh}_{K'}(G',X') \tag{33}$$

is a closed immersion. Consider the Hilbert Shimura datum (G,X) attached to a totally real field F and the corresponding Siegel datum (G',X') obtained by forgetting the action of F. Then

[41] Or “defined by” or “associated with”, but not “associated to”, which, strictly speaking, is not English. Careful writers distinguish “attach to” from “associate with”, and look with horror on “associate to”.

$\mathrm{Sh}_K(G,X)(\mathbb{C})$ classifies triples (A,i,η) with A a complex polarized abelian, i an action of $\mathcal{O}_F$, and η a level structure. The map (33) sends (A,i,η) to (A,η). Now, certainly, A_1 and A_2 can be isomorphic without (A_1,i_1) and (A_2,i_2) being isomorphic, but every isomorphism $A_1 \to A_2$ compatible with sufficiently high level structures will be compatible with the $\mathcal{O}_X$-structures. In the general situation, for a given K', there need not be a K for which (33) is a closed immersion, but there will be if K' is small enough.

The structure of a Shimura variety

By the structure of $\mathrm{Sh}(G,X)$, we 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,[42] 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 before, we let Z denote the centre of G and T the largest commutative quotient of G. There are homomorphisms $Z \hookrightarrow G \xrightarrow{\nu} T$, and we define

$$\begin{cases} T(\mathbb{R})^\dagger = \mathrm{Im}(Z(\mathbb{R}) \to T(\mathbb{R})), \\ T(\mathbb{Q})^\dagger = T(\mathbb{Q}) \cap T(\mathbb{R})^\dagger. \end{cases} \tag{34}$$

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 that 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

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

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 $G^{\mathrm{der}}(\mathbb{Q})$ containing $K \cap G^{\mathrm{der}}(\mathbb{Q})$.*

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

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

The theorem gives a diagram

$$\begin{array}{ccc} G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K & \hookleftarrow & \Gamma\backslash X^+ \\ \downarrow & & \downarrow \\ T(\mathbb{Q})^\dagger\backslash T(\mathbb{A}_f)/\nu(K) & \hookleftarrow & [1] \end{array} \tag{35}$$

[42]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.

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

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 over $\mathbb{Q}$.*

(a) *For every finite prime ℓ, the group $H^1(\mathbb{Q}_\ell, H)$ is trivial.*

(b) *The map $H^1(\mathbb{Q},H) \to \prod_{l\le\infty} H^1(\mathbb{Q}_l, H)$ is injective (Hasse principle).*

PROOF. (a) Platonov and Rapinchuk 1994, Theorem 6.4, p. 284.
(b) Platonov and Rapinchuk 1994, Theorem 6.6, p. 286. □

Both statements fail for groups that are not simply connected.

LEMMA 5.20. *Assume that 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:

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

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

$$g_{\mathbb{R}} \in G^{\mathrm{der}}(\mathbb{R})\cdot z \subset G^{\mathrm{der}}(\mathbb{R})\cdot Z(\mathbb{R}) \overset{5.2}{\subset} G(\mathbb{R})_+.$$

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 $t \in T(\mathbb{Q})^{\dagger}$. □

When G^{der} is simply connected, 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 $[\nu(g)] = [1]$, then $\nu(g) = \nu(q)\nu(k)$ for 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 that 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. It suffices to prove the following statements:

(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 (Milne 2017, 17.98), $H^1(\mathbb{F}_\ell, (\mathcal{G}'_\ell)_{\mathbb{F}_\ell}) = 0$, 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 7.31) 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 show that $T(\mathbb{Q})\backslash T(\mathbb{A}_f)/\nu(K)$ is finite. For a torus T over $\mathbb{Q}$, set

$$T(\mathbb{Z}_\ell) = \{a \in T(\mathbb{Q}_\ell) \mid \chi(a) \text{ is integral for all } \chi \in X^*(T)\} \text{ and}$$
$$T(\hat{\mathbb{Z}}) = \prod\nolimits_\ell T(\mathbb{Z}_\ell).$$

The ***class group*** of T is defined to be

$$H(T) = T(\mathbb{Q})\backslash T(\mathbb{R}) \times T(\mathbb{A}_f)/(T(\mathbb{R}) \times T(\hat{\mathbb{Z}})).$$

PROPOSITION 5.22. *The class group of every torus T over $\mathbb{Q}$ is finite.*

PROOF. If $T = (\mathbb{G}_m)_{F/\mathbb{Q}}$ with F a number field, then the class group of T is equal to the class group of F, and so the proposition follows from algebraic number theory. For the general case, see Ono 1959. □

It suffices to prove that $T(\mathbb{Q})\backslash T(\mathbb{A}_f)/\nu(K)$ is finite for K sufficiently small. Thus, we may suppose that $\nu(K) \subset T(\hat{\mathbb{Z}})$. As $T(\hat{\mathbb{Z}})$ is compact and $\nu(K)$ is open, $(T(\hat{\mathbb{Z}}){:}\,\nu(K))$ is finite. Therefore

$$T(\mathbb{Q})\backslash(T(\mathbb{R}) \times T(\mathbb{A}_f)/(T(\mathbb{R}) \times \nu(K))$$

is finite. As $T(\mathbb{Q})\backslash T(\mathbb{A}_f)/\nu(K)$ is a quotient of this group, it also is finite.

REMARK 5.23. Let (G,X) be a Shimura datum with G^{der} simply connected. We saw in the discussion preceding Lemma 5.20 that the fibre of $\mathrm{Sh}_K(G,X)$ over $[1] \in \pi_0(\mathrm{Sh}_K(G,X))$ is canonically isomorphic to $\Gamma\backslash X^+$, where Γ is the image of $K \cap G(\mathbb{Q})_+$ in $G^{\mathrm{ad}}(\mathbb{Q})^+$. When is this fibre equal $\Gamma\backslash X^+$ with $\Gamma = K \cap G^{\mathrm{der}}(\mathbb{Q})$? Equivalently, when does

$$\mathrm{Sh}_K(G,X)^\circ = \mathrm{Sh}^\circ_{K\cap G^{\mathrm{der}}(\mathbb{A}_f)}(G^{\mathrm{der}}, X^+)?$$

This is true whenever $Z' \overset{\mathrm{def}}{=} Z(G^{\mathrm{der}})$ satisfies the Hasse principle for H^1 (for then every element in $G(\mathbb{Q})_+ \cap K$ with K sufficiently small will lie in $G^{\mathrm{der}}(\mathbb{Q}) \cdot Z(\mathbb{Q})$; see the Addendum A). 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 $h{:}\,\mathbb{C}^\times \to T(\mathbb{R})$ defines a Shimura variety $\mathrm{Sh}(T,\{h\})$ — in this case the conditions SV1,2,3 are vacuous. For a compact open subgroup $K \subset T(\mathbb{A}_f)$,

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

(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), $Y\simeq T(\mathbb{Q})/T(\mathbb{Q})^{\dagger}$ and

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

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

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

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

For example, let $(G,X)=(\mathrm{GL}_2,\mathcal{H}_1^{\pm})$ and $K=K(N)$. Then $T=\mathbb{G}_m$ and $Y=\mathbb{R}/\mathbb{R}^{+}\simeq\{\pm1\}$ (see (34)). Thus

$$\pi_0(\mathrm{Sh}_K(G,X))=T(\mathbb{Q})\backslash\{\pm1\}\times\mathbb{A}_f^{\times}/(1+N\mathbb{A}_f^{\times})\simeq(\mathbb{Z}/N\mathbb{Z})^{\times}\simeq\mathrm{Gal}(\mathbb{Q}[\zeta_N]/\mathbb{Q}).$$

Additional axioms

Let (G,X) be a Shimura datum. The weight homomorphism w_X is a homomorphism $\mathbb{G}_{m\mathbb{R}}\to Z(G)^{\circ}_{\mathbb{R}}\subset G_{\mathbb{R}}$ over $\mathbb{R}$ of tori defined over $\mathbb{Q}$. It is therefore defined over $\mathbb{Q}^{\mathrm{a}}$. Some simplifications to the theory occur when some of the following conditions hold:

SV2*: for all $h\in X$, $\mathrm{ad}(h(i))$ is a Cartan involution on $G_{\mathbb{R}}/w_X(\mathbb{G}_m)$ (rather than $G^{\mathrm{ad}}_{\mathbb{R}}$);

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 p. 98 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 (p. 26). 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, and the natural field of definition of the Shimura variety is either a totally real or a CM-field (12.4 below).[43]

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 $\mathbb{R}$. We have

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

[43]In my view, the extra generality obtained by omitting SV6 is spurious, but Deligne disagrees with me.

Let X denote the $G(\mathbb{R})$-conjugacy class of h. Then (G,X) satisfies SV1 and SV2, and so it is a Shimura datum provided B splits (i.e., becomes isomorphic to $M_2(\mathbb{R})$) at at least one real prime of F. It then satisfies SV3 because G^{ad} is simple (as an algebraic group over $\mathbb{Q}$). Let $I = \mathrm{Hom}(F,\mathbb{Q}^{\mathrm{a}}) = \mathrm{Hom}(F,\mathbb{R})$, 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 $\mathbb{Q}^{\mathrm{a}}$ fixed by the automorphisms of $\mathbb{Q}^{\mathrm{a}}$ 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 $T(\mathbb{A}_f)$. The can be shown directly: the open subgroup $(1+4\mathbb{Z}_2)\times\prod_{p\text{ odd}}\mathbb{Z}_p^\times$ of $\mathbb{A}_f^\times$ intersects $\mathbb{Q}^\times$ in $\{1\}$.

(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 T is anisotropic if and only if $T(\mathbb{R})$ is compact. The subtorus T^a of T in 5.26 is the largest anisotropic subtorus of T (over $\mathbb{Q}$). 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 $\mathbb{R}$.

In particular, a Shimura datum (G,X) satisfies SV5 if and only if the largest anisotropic subtorus of $Z(G)$ remains anisotropic over $\mathbb{R}$.

A CM field L admits a nontrivial involution ι that becomes complex conjugation under every embedding of L into $\mathbb{C}$. Let T be a torus over $\mathbb{Q}$ that splits over a CM-field L. The subtorus

$$T_L^+ \overset{\text{def}}{=} \bigcap_{\iota\chi=-\chi} \operatorname{Ker}(\chi\colon T_L \to \mathbb{G}_m)$$

of T_L is defined over $\mathbb{Q}$ — it is the largest subtorus of T splitting over $\mathbb{R}$. Then T^+ splits over the maximal totally real subfield of L, and $T(\mathbb{Q})$ is discrete in $T(\mathbb{A}_f)$ if and only if T^+ splits 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

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

THEOREM 5.28. *Let (G,X) be a Shimura datum. Then*

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

When SV5 holds,

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

PROOF. The first equality can be proved by the same argument as 4.19,[44] and the second follows from the first. □

REMARK 5.29. Put $S_K = \mathrm{Sh}_K(G,X)$. 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'.

[44] The proof of Theorem 5.28 in Deligne 1979, 2.1.10, reads in its entirety:

> L'action de $G(\mathbb{Q})/Z(\mathbb{Q})$ sur $X \times \big(G(\mathbb{A}_f)/Z(\mathbb{Q})^-\big)$ est propre. Ceci permet le passage à la limite sur K.

Properness implies that the quotient of $X \times \big(G(\mathbb{A}_f)/Z(\mathbb{Q})^-\big)$ by $G(\mathbb{Q})/Z(\mathbb{Q})$ is Hausdorff (Bourbaki 1989, III 4.2), and hence Lemma 4.20 applies. Presumably, the action *is* proper, but I don't know a proof that the quotient is Hausdorff even in the easier case 4.19. The obvious argument doesn't seem to work.

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 $G(\mathbb{A}_f)$,

$$\mathrm{Sh}_K(G,X) = S/K \quad \text{(quotient of } S \text{ by the action of } K\text{)}$$

(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,

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

ASIDE. Does every arithmetic locally symmetric algebraic variety arise as a connected component of a Shimura variety? The answer is yes. More precisely, there is the following result (Milne 2013, 8.6): For every semisimple algebraic group H over $\mathbb{Q}$ and homomorphism $\bar{h}\colon \mathbb{S}/\mathbb{G}_m \to H^{\mathrm{ad}}_{\mathbb{R}}$ satisfying (SV1,2,3), there exists a reductive group G over $\mathbb{Q}$ with $G^{\mathrm{der}} = H$ and a homomorphism $h\colon \mathbb{S} \to G_{\mathbb{R}}$ lifting $\bar{h}$ and satisfying (SV1,2,2*,3,4,5,6).

ASIDE. Roughly speaking, there are two ways of describing reductive groups: concretely in terms algebras with involution and sesquilinear forms or more abstractly in terms of root data. Shimura always uses the first.[45] When Deligne was asked to report on Shimura's work in 1971, he did so in terms of abstract reductive groups, and he used root systems. For Shimura, the central object of study is the quotient $\Gamma\backslash X$ of a hermitian domain by a congruence subgroup of a semisimple group G over $\mathbb{Q}$. In order to define a canonical model of $\Gamma\backslash X$ over a number field, he needed to choose a reductive group with derived group G, but for him the reductive group was auxiliary. For Deligne, the reductive group is the starting point. His insight, that to define a Shimura variety, one needs only a reductive group G over $\mathbb{Q}$ and a homomorphism $h\colon \mathbb{S} \to G_{\mathbb{R}}$ satisfying certain axioms came as revelation to others trying to understand Shimura's work. For example, it helped Langlands in his effort to understand the zeta functions of Shimura varieties. In Shimura's description, the variety comes with much baggage, and it was not easy to discern what is essential.

NOTES. Here is a dictionary.[46]

This work:	SV1	SV2	SV3	SV4	SV2*
Deligne 1979:	2.1.1.1	2.1.1.2	2.1.1.3	2.1.1.4	2.1.1.5.

My axiom SV5 is weaker than (2.1.1.5) = SV2* but is exactly what is needed for most applications. Axiom SV6 is the weakest that keeps us in the realm of CM fields and their subfields.

[45] As far as I know, neither Shimura nor Weil ever use root systems. Langlands's familiarity with root systems and root data may have been one reason he was able to see so much further than everyone else in the 1960s and 1970s.

[46] As noted earlier, Deligne often writes 2.1.x when he means 2.1.1.x.

6 The Siegel modular variety

In this section, we study the most basic 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 giving 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 decomposition $V(\mathbb{C}) = V^{-1,0} \oplus V^{0,-1}$,

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

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

$$\psi(Ju,Jv) = \psi(u,v) \iff \psi(zu,zv) = |z|^2\psi(u,v) \text{ for all } z \in \mathbb{C}. \tag{36}$$

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

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

and

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

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 $(e_{\pm i})_{1\le i\le n}$ such that

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

This means that the matrix of ψ with respect to $(e_{\pm i})$ has ± 1 down the second diagonal, and zeros elsewhere:

$$\big(\psi(e_{\pm i},e_{\pm j})\big)_{1\le i,j\le g} = \begin{pmatrix} 0 & -I_g' \\ I_g' & 0 \end{pmatrix}, \quad I_g' = \begin{pmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{pmatrix}.$$

LEMMA 6.1. *Let W be a totally isotropic subspace of V. Then every 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. Certainly, the second statement is true when $n = 1$. We assume it inductively for spaces of dimension $\leq 2n-2$. Let W be totally isotropic, and let W' be a subspace of V such that $V = W^{\perp} \oplus W'$. Then $W^{\vee} \simeq V/W^{\perp} \simeq W'$ identifies W' with the dual of W. Let $e_1, \ldots, e_m$ be a basis for W, and let $e_{-1}, \ldots, e_{-m}$ be the dual basis in W'. Then $(e_{\pm i})_{1 \leq i \leq m}$ is a symplectic basis for $W \oplus W'$. By induction $(W \oplus W')^{\perp}$ has a symplectic basis $(e_{\pm i})_{m+1 \leq i \leq n}$, and then $(e_{\pm i})_{1 \leq i \leq n}$ is a symplectic basis for V. □

Thus, every maximal totally isotropic subspace of V has dimension n. Such subspaces are called ***lagrangians***.

Let (V, ψ) be a nonzero symplectic space, and 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

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

There is a homomorphism $\nu\colon \mathrm{GSp}(\psi) \to \mathbb{G}_m$ sending g to $\nu(g)$. The kernel of ν is the ***symplectic group*** $\mathrm{Sp}(\psi)$, which is the derived group of $\mathrm{GSp}(\psi)$. We have a diagram

$$\begin{tikzcd}
 & \mathrm{Sp}(\psi) \arrow[d] \arrow[dr] & \\
\mathbb{G}_m \arrow[r] \arrow[dr] & \mathrm{GSp}(\psi) \arrow[r, "\mathrm{ad}"] \arrow[d, "\nu"] & \mathrm{GSp}(\psi)^{\mathrm{ad}} \\
 & \mathbb{G}_m &
\end{tikzcd}$$

in which the column and row are short exact sequences, and the diagonal maps are isogenies with kernel the centre $\mathbb{G}_m \cap \mathrm{Sp}(\psi) = \mu_2$ of $\mathrm{Sp}(\psi)$.

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 every automorphism of V, 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 $G = G(\psi) = \mathrm{GSp}(\psi)$ and $S = S(\psi) = \mathrm{Sp}(\psi) = G^{\mathrm{der}}$.

Let J be a complex structure on $V(\mathbb{R})$ such that $\psi(Ju, Jv) = \psi(u, v)$. Then $J \in S(\mathbb{R})$. For all $z \in \mathbb{C}^{\times}$, $h_J(z)$ lies $G(\mathbb{R})$ by (36), and it lies in $S(\mathbb{R})$ if $|z| = 1$. We say that J is ***positive*** (resp. ***negative***) if $\psi_J(u, v) \overset{\text{def}}{=} \psi(u, Jv)$ is positive-definite (resp. negative-definite).

Let X^+ (resp. X^-) denote the set of positive (resp. negative) complex structures on $V(\mathbb{R})$ such that $\psi(Ju, Jv) = \psi(u, v)$ for all $u, v \in V$, 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 because the element $g: e_{\pm i} \mapsto e_{\mp i}$ of $G(\mathbb{R})$ has $\nu(g) = -1$ and it interchanges X^+ and 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 the map $J \mapsto h_J$ identifies X with a $G(\mathbb{R})$-conjugacy class of homomorphisms $h: \mathbb{C}^\times \to G(\mathbb{R})$. We often denote X by $X(\psi)$ and X^+ by $X(\psi)^+$.

EXERCISE 6.2. (a) Show that for any $h \in X(\psi)$, $\nu(h(z)) = z\bar{z}$. [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) Let $\dim V = 2g$. 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 PAIR $(G(\psi), X(\psi))$ SATISFIES THE AXIOMS SV1–SV6.

We let $\mathrm{GL}(V)$ act on $\mathrm{Hom}(V, V)$ according to the rule

$$(\alpha f)(v) = \alpha(f(\alpha^{-1}v)), \quad \alpha \in \mathrm{GL}(V), \quad f \in \mathrm{Hom}(V, V), \quad v \in V.$$

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 $\bar{z}$ respectively. Then

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

The Lie algebra of G is the subspace

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

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: The symplectic group is simple over every algebraically closed field, because its root system is indecomposable. Therefore, G^{ad} is $\mathbb{Q}$-simple. As $G^{\mathrm{ad}}(\mathbb{R})$ is not compact, we see that SV3 holds.

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 $\mathbb{Q}$.
This is the ***Siegel Shimura datum***.

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 $V(\mathbb{A}_f) = V \otimes_{\mathbb{Q}} \mathbb{A}_f$. 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 $\mathcal{H}_K$ denote the set of triples $((W,h),s,\eta K)$, where

- ◇ (W,h) is a rational Hodge structure of type $(-1,0),(0,-1)$;
- ◇ s or $-s$ is a polarization for (W,h);
- ◇ ηK is a K-orbit of $\mathbb{A}_f$-linear isomorphisms $V(\mathbb{A}_f) \to W(\mathbb{A}_f)$ under which ψ corresponds to an $\mathbb{A}_f^{\times}$-multiple of s.[47]

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 sending[48] s to cs' for some $c \in \mathbb{Q}^{\times}$ and such that $b \circ \eta = \eta' \mod K$.

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 $a\colon W \to V$ under which ψ corresponds to a $\mathbb{Q}^{\times}$-multiple of s. Then
$$ah \overset{\text{def}}{=} (z \mapsto a \circ h(z) \circ a^{-1})$$
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 $\mathrm{Sh}_K(\mathbb{C})$ classifies the elements of $\mathcal{H}_K$ 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. We first check that the the map sends isomorphic triples to the same class. Suppose that $b\colon (W,h) \to (W',h')$ is an isomorphism sending s to a $\mathbb{Q}^{\times}$-multiple of t' and that $b \circ \eta = \eta' \circ k$ for some $k \in K$. Choose an isomorphism $a'\colon W' \to V$ sending s' to a $\mathbb{Q}^{\times}$-multiple of ψ, and let $a = a' \circ b$. Then $(ah, a\circ\eta) = (a'h, a'\circ\eta'\circ k)$.

[47] The notation ηK is unfortunate since it suggests a K-orbit with a distinguished element. We mean only a K-orbit of isomorphisms.

[48] An isomorphism $b\colon W \to W'$ of vector spaces over $\mathbb{Q}$ defines an isomorphism $b'\colon \mathrm{Hom}(W \otimes W, \mathbb{Q}) \to \mathrm{Hom}(W' \otimes W', \mathbb{Q})$. When $b'(s) = s'$ we say that b sends s to s'.

We next check that two triples are isomorphic if they map to the same class. Let $(W\ldots)$ and $(W'\ldots)$ map to the same class. Choose isomorphisms $a\colon V \to W$ and $a'\colon V \to W'$ sending ψ to multiples of s and s'. We are given that $(ah, a\circ\eta) = (qa'h, q\circ a'\circ\eta\circ k)$ for some q and k. After replacing a' with $q\circ a'$, we may suppose that $(ah, a\circ\eta) = (a'h, a'\circ\eta\circ k)$. Then $b = a'\circ a^{-1}$ is an isomorphism $((W,h),\ldots) \to ((W',h'),\ldots)$.

Finally, 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.[49] 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 so $\pi_1(M,0) = \Lambda$ (Hatcher 2002, 1.40). Therefore, (ibid. 2A.1)

$$H_1(M,\mathbb{Z}) \simeq \Lambda \tag{37}$$

and (Greenberg 1967, 23.14)

$$H^1(M,\mathbb{Z}) \simeq \operatorname{Hom}(\Lambda,\mathbb{Z}). \tag{38}$$

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

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

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

$$(f_1\wedge\cdots\wedge f_n, v_1\otimes\cdots\otimes v_n) \mapsto \det(f_i(v_j))\colon \textstyle\bigwedge^n \Lambda^\vee \times \bigwedge^n \Lambda \to \mathbb{Z}$$

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

$$\operatorname{Hom}(\textstyle\bigwedge^n \Lambda,\mathbb{Z}) \simeq \bigwedge^n \operatorname{Hom}(\Lambda,\mathbb{Z}).$$

From (38) we see that it suffices to show that cup-product defines an isomorphism

$$\textstyle\bigwedge^n H^1(M,\mathbb{Z}) \to H^n(M,Z). \tag{39}$$

Let $\mathcal{T}$ be the class of topological manifolds M such that $H^1(M,\mathbb{Z})$ is a free $\mathbb{Z}$-module of finite rank and the maps (39) 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 $\mathcal{T}$. □

[49] Let $\pi\colon\mathbb{C}^n \to \mathbb{C}^n/\Lambda$ be the quotient map. A subset U of $\mathbb{C}^n/\Lambda$ is open if and only if $\pi^{-1}(U)$ is open and a function f on an open subset U of $\mathbb{C}^n/\Lambda$ is holomorphic if and only if $f\circ\pi$ is holomorphic on $\pi^{-1}(U)$.

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 $\tilde{\varphi}(0) = 0$:

$$\begin{array}{ccc} \mathbb{C}^n & \xrightarrow{\tilde{\varphi}} & \mathbb{C}^{n'} \\ \downarrow & & \downarrow \\ \mathbb{C}^n/\Lambda & \xrightarrow{\varphi} & \mathbb{C}^{n'}/\Lambda'. \end{array}$$

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 periodic function with respect to Λ, 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

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

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 a field k: a regular map $\varphi\colon A \to B$ of abelian varieties such that $\varphi(0) = 0$ is a homomorphism (Milne 1986, 3.6). For example, the map sending an element to its inverse is a homomorphism, which implies that the group law on A 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 $\mathbb{R} \otimes \Lambda \simeq \mathbb{C}^n$ 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 the following "period relations" hold:

$$\begin{cases} \psi_{\mathbb{R}}(Ju, Jv) = \psi_{\mathbb{R}}(u, v) & \text{for all } u, v \in V, \text{ and} \\ \psi_{\mathbb{R}}(u, Ju) > 0 & \text{for all } u \neq 0. \end{cases} \tag{40}$$

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. Alternatively, one can apply the Kodaira embedding theorem (Voisin 2002, Théorème 7.11 and Section 7.2.2). $\square$

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 $\mathrm{Tgt}_0(A)/\Lambda \simeq A(\mathbb{C})$ of complex manifolds (Mumford 1970, p. 2).

Let $M = \mathbb{C}^n/\Lambda$ be a complex torus. The complex structure on $\Lambda \otimes_{\mathbb{Z}} \mathbb{R}$ defined by the isomorphism $\Lambda \otimes_{\mathbb{Z}} \mathbb{R} \simeq \mathbb{C}^n$ endows $\Lambda \simeq H_1(M,\mathbb{Z})$ with an integral Hodge structure of weight -1 (see p. 67). Note that a Riemann form for M is nothing but a polarization of the integral Hodge structure Λ.

THEOREM 6.8 (RIEMANN'S THEOREM). [50] *The functor $A \rightsquigarrow H_1(A,\mathbb{Z})$ 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

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

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 $A \rightsquigarrow H_1(A,\mathbb{Q})$ 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. □

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

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

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

$$\mathrm{Tgt}_0(A) \simeq H_1(A,\mathbb{C})/F^0 \tag{41}$$

(isomorphism of complex vector spaces).

[50] 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.

A modular description of the points of the Siegel variety

Let A be an abelian variety over $\mathbb{C}$. We make the following definitions:

$$T_f(A) \stackrel{\text{def}}{=} H_1(A,\mathbb{Z}) \otimes_{\mathbb{Z}} \hat{\mathbb{Z}} \simeq \varprojlim_n H_1(A,\mathbb{Z})/nH_1(A,\mathbb{Z})$$

$$V_f(A) \stackrel{\text{def}}{=} H_1(A,\mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{A}_f \simeq T_f(A) \otimes_{\mathbb{Z}} \mathbb{Q}.$$

Thus $T_f(A)$ is a free $\hat{\mathbb{Z}}$-module of rank $2\dim A$. Let $A(\mathbb{C})_n = \mathrm{Ker}(n\colon A(\mathbb{C}) \to A(\mathbb{C}))$. If $A(\mathbb{C}) = \mathbb{C}^g/\Lambda$, then $H_1(A,\mathbb{Z}) \simeq \Lambda$ and

$$A(\mathbb{C})_n \simeq n^{-1}\Lambda/\Lambda \simeq \Lambda/n\Lambda.$$

Therefore,

$$T_f(A) \simeq \varprojlim_n A(\mathbb{C})_n.$$

More generally, for an abelian variety A over a field k of characteristic zero, we define

$$T_f(A) = \varprojlim_n A(k^{\mathrm{a}})_n$$

$$V_f(A) = T_f(A) \otimes_{\mathbb{Z}} \mathbb{Q}.$$

Let (V,ψ) be a symplectic space over $\mathbb{Q}$. Let $\mathcal{M}_K$ denote the set of triples $(A,s,\eta K)$, where

- ◇ 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 $V(\mathbb{A}_f) \to V_f(A)$ under which ψ corresponds 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* $\mathrm{Sh}_K(\mathbb{C})$ *classifies the elements* $(A,s,\eta K)$ *of* $\mathcal{M}_K$ *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 Corollary 6.9 with Proposition 6.3. □

ASIDE. As befits a Summer School on the Langlands program, my lectures were relentlessly adèlic. In the this aside, we restate Theorem 6.11 in more down-to-earth terms.

Let (G,X) be a Shimura datum, and let K be a compact open subgroup of $G(\mathbb{A}_f)$. When G^{der} is simply connected, we have the fundamental diagram (35),

$$\begin{tikzcd} \mathrm{Sh}_K(G,X)(\mathbb{C}) = G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K \arrow[d] & \Gamma\backslash X^+ \arrow[l, hook'] \arrow[d] \\ \mathrm{Sh}_{\nu(K)}(T,Y)(\mathbb{C}) = T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K) & {[1].} \arrow[l, hook'] \end{tikzcd} \tag{42}$$

If the centre of G^{der} satisfies the Hasse principle for H^1, then $\Gamma = K \cap G^{\mathrm{der}}(\mathbb{Q})$ (5.23).

Let (G, X) be the Shimura datum attached to a symplectic space (V, ψ) over $\mathbb{Q}$, and let $S = G^{\mathrm{der}} = \mathrm{Sp}(\psi)$. We suppose that there exists a $\mathbb{Z}$-lattice $V(\mathbb{Z})$ in V such that ψ restricts to a pairing $V(\mathbb{Z}) \times V(\mathbb{Z}) \to \mathbb{Z}$ with discriminant ± 1. For $N \geq 3$, let

$$K(N) = \{g \in G(\mathbb{A}_f) \mid g \text{ preserves } V(\hat{\mathbb{Z}}) \text{ and acts as 1 on } V(\hat{\mathbb{Z}})/NV(\hat{\mathbb{Z}})\}.$$

Let $S(\mathbb{Z})$ be the set of g in $S(\mathbb{Q})$ such that $gV(\mathbb{Z}) = V(\mathbb{Z})$. Then $K(N) \cap S(\mathbb{Q})$ is

$$\Gamma(N) = \{g \in S(\mathbb{Z}) \mid g \text{ acts 1 on } V(\mathbb{Z})/NV(\mathbb{Z})\}.$$

Let us write $V(\mathbb{Z}/N\mathbb{Z})$ for

$$V(\mathbb{Z})/NV(\mathbb{Z}) = V(\hat{\mathbb{Z}})/NV(\hat{\mathbb{Z}}).$$

It is a free $\mathbb{Z}/N\mathbb{Z}$-module of rank $\dim(V)$ with a perfect pairing ψ_N. As in the case $\dim V = 2$ (see p. 63), $\pi_0(\mathrm{Sh}_{K(N)}(G, X)) \simeq (\mathbb{Z}/N\mathbb{Z})^\times$, and so the diagram (42) becomes

$$\begin{tikzcd} \mathrm{Sh}_{K(N)}(\mathbb{C}) \arrow[d] & \Gamma(N)\backslash X^+ \arrow[l, hook'] \arrow[d] \\ (\mathbb{Z}/N\mathbb{Z})^\times & {[1].} \arrow[l, hook'] \end{tikzcd} \tag{43}$$

Let (A, λ) be an abelian variety of dimension $\frac{1}{2}\dim(V)$ over $\mathbb{C}$ with a principal polarization λ. From λ we get a perfect alternating pairing

$$e_N^\lambda : A(\mathbb{C})_N \times A(\mathbb{C})_N \to \mu_N$$

(see, e.g., Milne 1986, §16). We fix a choice of $i = \sqrt{-1}$ in $\mathbb{C}$. This gives us an isomorphism

$$[n] \mapsto e^{2\pi i n/N} : \mathbb{Z}/N\mathbb{Z} \to \mu_N,$$

and so e_N^λ becomes a pairing to $\mathbb{Z}/N\mathbb{Z}$. By a ***level-N structure*** on A, we mean an isomorphism

$$\eta : V(\mathbb{Z}/N\mathbb{Z}) \to A(\mathbb{C})_N$$

under which ψ_N corresponds to a $(\mathbb{Z}/N\mathbb{Z})^\times$ multiple of e_N^λ. The set $\mathrm{Sh}_K(\mathbb{C})$ classifies the isomorphism classes of pairs $((A, \lambda), \eta_N)$ consisting of a principally polarized abelian variety and a level N-structure. The map to $(\mathbb{Z}/N\mathbb{Z})^\times$ is the obvious one, and the fibre over [1] consists of the pairs $((A, \lambda), \eta_N)$ such that, under η_N, the pairings ψ_N and e_N^λ correspond exactly.

ASIDE. Traditionally, in studying Shimura varieties, we take K to be "sufficiently small". But then the varieties we get are of logarithmic general type (see the proof of 3.21), and hence less interesting to algebraic geometers. As in the above aside, take $V(\mathbb{Z})$ be a lattice on which ψ gives a perfect pairing, and let $\Gamma(1) = \{g \in G^{\mathrm{der}}(\mathbb{Q}) \mid gV(\mathbb{Z}) = V(\mathbb{Z})\}$. Let $n = \dim(V)/2$. Then $\mathcal{A}_n \stackrel{\mathrm{def}}{=} \Gamma(1)\backslash X^+$ is the Siegel modular variety of degree n. It is the coarse moduli variety of principally polarized abelian varieties of dimension n. The algebraic geometers have shown that $\mathcal{A}_n$ is unirational for $n \leq 5$ and of general type for $n \geq 7$ (the case $n = 6$ remains open).

7 Shimura varieties of Hodge type

In this section, we study an important generalization of Siegel modular varieties.

DEFINITION 7.1. A Shimura datum (G,X) is of ***Hodge type***[51] 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., $g\cdot x = \nu(g)^r\cdot x$ for $g\in G(\mathbb{Q})$ and $x\in\mathbb{Q}(r)$. For each $h\in X$, $(\mathbb{Q}(r),\nu\circ h)$ 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. Apply Chevalley's theorem, as in the proof of Proposition 2.1, to find tensors t_i in $V^{\otimes r_i}\otimes V^{\vee\otimes s_i}$ such that G is the subgroup of $G(\psi)$ fixing the t_i. But ψ defines an isomorphism $V\simeq V^\vee\otimes\mathbb{Q}(1)$, and so

$$V^{\otimes r_i}\otimes V^{\vee\otimes s_i}\simeq V^{\vee\otimes(r_i+s_i)}\otimes\mathbb{Q}(r_i)\simeq \mathrm{Hom}(V^{\otimes(r_i+s_i)},\mathbb{Q}(r_i)). \qquad \square$$

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 $\mathcal{H}_K$ denote the set of triples $((W,h),(s_i)_{0\le i\le n},\eta K)$, where

- ◇ (W,h) is a rational Hodge structure of type $(-1,0),(0,-1)$,
- ◇ s_0 or $-s_0$ 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 $V(\mathbb{A}_f)\to W(\mathbb{A}_f)$ under which ψ corresponds to an $\mathbb{A}_f^\times$-multiple of s_0 and t_i corresponds to s_i for $i=1,\ldots,n$,

satisfying the following condition:

(*) there exists an isomorphism $a\colon W\to V$ under which s_0 corresponds to a $\mathbb{Q}^\times$-multiple of ψ, s_i corresponds to t_i for $i=1,\ldots,n$, and h corresponds to 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', the tensor s_i to s_i' for every i, and ηK to $\eta' K$.

PROPOSITION 7.3. *The set $\mathrm{Sh}_K(\mathbb{C})$ classifies the elements of $\mathcal{H}_K$ modulo isomorphism.*

PROOF. Let $((W,h),(s_i),\eta K)\in\mathcal{H}_K$. Choose an isomorphism $a\colon W\to V$ as in (*), and consider the pair $(ah,a\circ\eta)$, where $(ah)(z)=a\circ h(z)\circ a^{-1}$. 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 by its action on $s_0,\ldots,s_n$ up to composition with an

[51] As far as I know, the term was introduced in Milne 1990, II, §3.

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 the map $(W,\ldots)\mapsto [ah, a\circ\eta]_K$ gives a bijection from the set of isomorphism classes of elements of $\mathcal{H}_K$ onto $\mathrm{Sh}_K(G,X)(\mathbb{C})$ is straightforward (as for 6.3). $\square$

Let $t\colon V\times\cdots\times V\to\mathbb{Q}(r)$ (m-copies of V) be a multilinear map 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 $W=H_1(A,\mathbb{Q})$. Then (see 6.4)

$$H^m(A,\mathbb{Q})\simeq \mathrm{Hom}(\textstyle\bigwedge^m W,\mathbb{Q}).$$

We say that $t\in H^{2r}(A,\mathbb{Q})$ is a ***Hodge tensor for*** A if the corresponding map

$$W^{\otimes 2r}\to\textstyle\bigwedge^{2r}W\to\mathbb{Q}(r)$$

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 $(A,(s_i)_{0\le i\le n},\eta K)$ in which

- $\diamond$ A is a complex abelian variety,
- $\diamond$ s_0 or $-s_0$ is a polarization for the rational Hodge structure $H_1(A,\mathbb{Q})$,
- $\diamond$ $s_1,\ldots,s_n$ are Hodge tensors for A or its powers, and
- $\diamond$ 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 $i\ge 1$, and h to an element of X.

An isomorphism from one triple $(A,(s_i)_i,\eta K)$ to a second $(A',(s'_i)_i,\eta' K)$ 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* $\mathrm{Sh}_K(\mathbb{C})$ *classifies the elements of* $\mathcal{M}_K$ *modulo isomorphism.*

PROOF. Combine Propositions 7.3 and 6.9. $\square$

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, it is sometimes possible to replace (**) by a simpler trace condition.

REMARK 7.5. If we let $A(\mathbb{C})=\mathbb{C}^g/\Lambda$, then (see 6.4),

$$H^m(A,\mathbb{Q})\simeq\mathrm{Hom}(\textstyle\bigwedge^m\Lambda,\mathbb{Q})$$

Now $\Lambda\otimes\mathbb{C}\simeq T\oplus\bar{T}$, where $T=\mathrm{Tgt}_0(A)$. Therefore,

$$H^m(A,\mathbb{C})\simeq\mathrm{Hom}(\textstyle\bigwedge^m(\Lambda\otimes\mathbb{C}),\mathbb{C})\simeq\mathrm{Hom}(\bigoplus\limits_{p+q=m}\bigwedge^p T\otimes\bigwedge^q\bar{T},\mathbb{C})\simeq\bigoplus\limits_{p+q=m}H^{p,q},$$

where

$$H^{p,q} = \mathrm{Hom}(\bigwedge^p T \otimes \bigwedge^q \bar{T}, \mathbb{C}).$$

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 cohomology class of every algebraic cycle on A is a Hodge class, and the Hodge conjecture predicts space of Hodge tensors is the $\mathbb{Q}$-span of the set of algebraic classes. For $r = 1$, this is easy to prove. The exponential sequence

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathcal{O}_A \xrightarrow{z \mapsto \exp(2\pi i z)} \mathcal{O}_A^\times \longrightarrow 0$$

gives a cohomology sequence

$$H^1(A, \mathcal{O}_A^\times) \to H^2(A,\mathbb{Z}) \to H^2(A, \mathcal{O}_A). \tag{44}$$

The cohomology group $H^1(A, \mathcal{O}_A^\times)$ classifies the divisors on A modulo linear equivalence, i.e., $H^1(A, \mathcal{O}_A^\times) \simeq \mathrm{Pic}(A)$, 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

$$\mathrm{Im}(\mathrm{Pic}(A)) = H^2(A,\mathbb{Z}) \cap H^{1,1}.$$

Thus, the Hodge conjecture holds for $r = 1$ even for integral cohomology groups. This case of the conjecture was orginally proved by Lefschetz, and the result is often called the Lefschetz $(1,1)$-theorem.

From (44) we get an injective homomorphism

$$\mathrm{NS}(A) \to H^2(A,\mathbb{Z}) \simeq \mathrm{Hom}(\bigwedge^2 H_1(A,\mathbb{Z}), \mathbb{Z}),$$

where $\mathrm{NS}(A)$ is the Nèron-Severi group of A. By a ***polarization*** of A we mean an element of $\mathrm{NS}(A)$ mapping to a polarization of $H_1(A,\mathbb{Z})$.

8 PEL Shimura varieties

In this section, we construct the Shimura varieties classifying polarized abelian varieties with an action of the ring of integers in a fixed semisimple $\mathbb{Q}$-algebra B and a level structure. The construction is complicated by the fact that the algebraic group defining the Shimura variety is attached, not to B, but to the centralizer of B in some endomorphism algebra. Throughout this section, k is a field of characteristic zero, and bilinear forms are always nondegenerate.

Algebras with involution

By a ***k-algebra*** in this section, we mean a ring containing k in its centre and of finite dimension as a k-vector space.

A k-algebra A is ***simple*** if 0 and A are its only two-sided ideals. Every matrix algebra $M_n(D)$, $n \geq 1$, over a division k-algebra D is simple, and a theorem of Wedderburn says that all simple k-algebras are of this form (CFT, IV 1.15). When k is algebraically closed, the only division k-algebra is k itself, and so the only simple k-algebras are the matrix algebras. A simple k-algebra B has only one simple module M up to isomorphism, and every B-module is isomorphic to a direct sum of copies of M (CFT, IV 1.18). For example, D^n is the only simple $M_n(D)$-module up to isomorphism.

A k-algebra B is ***semisimple*** if every B-module is semisimple, i.e., a direct sum of simple modules. For example, a simple k-algebra is semisimple. A semisimple k-algebra B has only finitely many minimal two-sided ideals, $B_1, \ldots, B_r$, each B_i is a simple k-algebra, and $B \simeq B_1 \times \cdots \times B_m$. 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. The ***trace map*** of a B-module M is the k-linear map

$$b \mapsto \mathrm{Tr}_k(b|M) \colon B \to k.$$

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 $B_1, \ldots, B_m$ 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_j. We have to show that the trace map determines the multiplicities r_j. But for $e_i = (\underset{1}{0}, \ldots, 0, \underset{i}{1}, 0, \ldots)$,

$$\mathrm{Tr}_k\left(e_i \,\middle|\, \textstyle\sum_j r_j M_j\right) = r_i \dim_k M_i\,. \qquad \square$$

REMARK 8.2. When k has characteristic $p \neq 0$, the proposition fails because the trace map of pM is zero.

An ***involution*** of a k-algebra B is a k-linear bijection $b \mapsto b^* \colon B \to B$ such that $(ab)^* = b^* a^*$ and $b^{**} = b$ for all $a, b \in B$. Note that 1^* is then an identity element and so $1^* = 1$; it follows that $c^* = c$ for all $c \in k$. An involution of B maps the centre of B into itself. When it fixes the elements of the centre, it is said to be of the ***first kind***; otherwise it is of the ***second kind.***

PROPOSITION 8.3. *Let $(B,*)$ be a semisimple k-algebra with involution. If the field k is algebraically closed, 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) $M_n(k), \quad b^* = b^t$ *(orthogonal type);*

(BD) $M_n(k), \quad b^* = J \cdot b^t \cdot J^{-1}, \quad J = \left(\begin{smallmatrix} 0 & -I \\ I & 0 \end{smallmatrix}\right)$ *(symplectic type).*

PROOF. In every decomposition $B = B_1 \times \cdots \times B_r$ of B into a product of simple k-algebras, the B_i are the minimal two-sided ideals of B, and so the set $\{B_1, \ldots, B_r\}$ is uniquely determined by B. On applying $*$, we get a decomposition $B = B_1^* \times \cdots \times B_r^*$ with $(B_1^*, \ldots, B_r^*)$ a permutation of $(B_1, \ldots, B_r)$. It follows that B is a product of semisimple k-algebras with involution, each of which is either (a) simple or (b) the product of two simple algebras interchanged by $*$.

Let $(B,*)$ be as in (a). As k is algebraically closed, B is isomorphic to $M_n(k)$ for some n, and the theorem of Skolem and Noether [52] says 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. Replacing the isomorphism $B \to M_n(k)$ with its composite with $x \mapsto gxg^{-1}$ for some $g \in \mathrm{GL}_n(k)$ changes u to gug^t, and so we may suppose that u is the identity matrix I or the matrix J. Hence $(B,*)$ is of type (C) or (BD).

Let $(B,*)$ be as in (b), say $B = B_1 \times B_2$. Then $*$ is an isomorphism of the opposite B_1^{opp} of B_1 onto B_2. Because k is algebraically closed, it follows 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 $a \leftrightarrow a^t : M_n(k)^{\mathrm{opp}} \simeq M_n(k)$ to see that $(B,*)$ is of type (A). □

Note that the involution $*$ is of the second kind in case (A) and of the first kind in cases (C) and (BD).

Let W be a (finite-dimensional) vector space over k, and let $\phi : W \times W \to k$ be a (nondegenerate) bilinear form on W. For $\alpha \in \mathrm{End}_k(W)$, define α^* to be the endomorphism such that

$$\phi(\alpha^*(v), w) = \phi(v, \alpha(w)), \qquad \text{all } v, w \in W.$$

The map $\alpha \mapsto \alpha^*$ is an involution of the k-algebra $\mathrm{End}_k(W)$ if (and only if) ϕ is symmetric or skew-symmetric. We call it the ***adjoint involution*** of ϕ.

We can now restate Proposition 8.3 as follows.

PROPOSITION 8.4. *Let $(B,*)$ be a semisimple k-algebra with involution. If the field k is algebraically closed and the only elements of the centre of B fixed by $*$ are those in k, then $(B,*)$ is isomorphic to one of the following:*

(A) $(\mathrm{End}_k(W) \times \mathrm{End}_k(W^\vee), *)$ *with* $(a,b)^* = (b^\vee, a^\vee)$;

(C) $(\mathrm{End}_k(W), *)$ *with $*$ the adjoint involution of a symmetric bilinear form on W;*

(BD) $(\mathrm{End}_k(W), *)$ *with $*$ the adjoint involution of an alternating bilinear form on W.*

The labelling is explained in the next section, where we attach an algebraic group matching the label to a symplectic module over $(B,*)$.

[52] Let $f, g : A \rightrightarrows B$ be homomorphisms from a k-algebra A to a k-algebra B. If A is simple and B is central simple over k, then there exists an invertible element $b \in B$ such that $f(a) = b \cdot g(a) \cdot b^{-1}$ for all $a \in A$ (CFT, IV 2.10).

Symplectic modules and the associated algebraic groups

Let B be a k-algebra with involution $*$. A ***symplectic*** $(B,*)$***-module*** is a B-module V equipped with a skew-symmetric k-bilinear form $\psi\colon V\times V\to k$ such that

$$\psi(b^{*}u,v)=\psi(u,bv) \text{ for all } b\in B, \text{ and } u,v\in V. \tag{45}$$

In general, a k-bilinear form $\psi\colon V\times V\to k$ satisfying (45) is said to be ***balanced***.

Let B be a semisimple k-algebra with centre F. Then F is a product of fields $F=\prod F_i$ and correspondingly $B=\prod B_i$ with $B_i=B\otimes_F F_i$ a central simple algebra over F_i. Assume that B is free as an F-module. This means that the degree $[B_i\colon F_i]=n^2$ is independent of i. For a B-module (finite-dimensional as a k-module), there is a ***reduced determinant map***

$$\det\colon \mathrm{End}_B(V)\to F.$$

If there exists an isomorphism $u\colon M_n(F)\to B$, then

$$\det(g)=\det(g|u(E_{1,1})\cdot V).$$

For example, if $B=F$ acting on $V=F^m$, then this is the usual determinant $\det\colon M_m(F)\to F$, and if $B=M_m(F)$ acting on $V=F^m$, then it is the identity map $F^\times\to F^\times$ (which is an mth root of the usual determinant of F acting on F^m). In the general case we may suppose that F is a field; then for some finite Galois extension E of F, there exists an isomorphism $u\colon M_n(E)\to B\otimes_F E$, and hence a reduced determinant map $\det\colon \mathrm{End}_{B\otimes E}(V\otimes E)\to E$; if $g\in \mathrm{End}_B(V)\subset \mathrm{End}_{B\otimes E}(V\otimes E)$, then $\det(g)$ is fixed by $\mathrm{Gal}(E/F)$, and so lies in F. The map det we obtain is independent of all the choices. When B is a central simple algebra over k acting on a simple B-module, the determinant map is the reduced norm.

ASIDE. It will be useful to review the relation between (skew-)hermitian forms and involutions (Knus et al. 1998, I, §4). Let $(A,*)$ be a central simple algebra with involution over a field F. A ***hermitian*** (resp. ***skew-hermitian***) form on a (left) A-module is V is a bi-additive map $\phi\colon V\times V\to A$ such that $\phi(au,bv)=a\phi(u,v)b^*$ and $\phi(v,u)=\phi(u,v)^*$ (resp. $\phi(v,u)=-\phi(u,v)^*$) for all $a,b\in A$ and $u,v\in V$. As in the bilinear case, a (nondegenerate) hermitian or skew-hermitian form ϕ on V defines an adjoint involution $*_\phi$ on $B\overset{\text{def}}{=}\mathrm{End}_A(V)$ by $\phi(\alpha^{*_\phi}u,v)=\phi(u,\alpha v)$.

(a) When $*$ is of the first kind, this gives a one-to-one correspondence between the involutions of the first kind on B and the forms ϕ on V, hermitian or skew-hermitian, up to a factor in $F^\times$. If ϕ is hermitian, then $*$ and $*_\phi$ have the same type, and if ϕ is skew-hermitian then they have the opposite type (e.g., if $*$ on A is of type (C) then $*_\phi$ on B is of type (BD)).

(b) When $*$ is of the second kind, this gives a one-to-one correspondence between the extensions of $*|F$ to B and the hermitian forms on V up to a factor in $F^\times$ fixed by $*$.

Suppose that $*$ is of the second kind. Then F is of degree 2 over the fixed field F_0 of $*$. Choose an element f of $F\smallsetminus F_0$ whose square is in F_0. Then $f^*=-f$, and a pairing ϕ is hermitian (resp. skew-hermitian) if and only if $f\phi$ is skew-hermitian (resp. hermitian). Thus (b) also holds with "skew-hermitian" for "hermitian".

Let $(B,*)$ be a semisimple k-algebra with involution $*$. Let F denote the centre of B and F_0 the subalgebra of invariants of $*$ in F. We say that $(B,*)$ has type (A), (C), or (BD) if $(B\otimes_{F_0,\rho}k^{\mathrm{a}},*)$ has that type for all k-homomorphisms $\rho\colon F_0\to k^{\mathrm{a}}$. This will always be the case if F_0 is a field. Assume that B is free as an F-module. Then F is free of rank 2 over F_0 in case (A) and equals F_0 in the cases (C) and (BD). We put $[B\colon F]=n^2$ and $[F_0\colon k]=g$.

Let (V,ψ) be a symplectic $(B,*)$-module with V free over F. The reduced dimension of V is

$$m = \frac{\dim_F(V)}{[B{:}F]^{1/2}} = \frac{\dim_F(V)}{n}.$$

Let G_1 and G be the algebraic subgroups of $\mathrm{GL}_B(V)$ such that

$$\begin{aligned} G_1(k) &= \{g \in \mathrm{GL}_B(V) \mid \psi(gx,gy) = \psi(\mu(g)x,y) \text{ for some } \mu(g) \in F_0^{\times}\} \\ G(k) &= \{g \in \mathrm{GL}_B(V) \mid \psi(gx,gy) = \mu(g)\cdot\psi(x,y) \text{ for some } \mu(g) \in k^{\times}\}. \end{aligned}$$

There is a homomorphism of algebraic groups $\mu{:}\, G_1 \to (\mathbb{G}_m)_{F_0/k}$, and G is the inverse image of $\mathbb{G}_m \subset (\mathbb{G}_m)_{F_0/k}$. We put

$$G' = \mathrm{Ker}(\mu) \cap \mathrm{Ker}(\det), \quad T_1 = G_1/G', \quad T = G/G'.$$

The following diagrams summarize the situation:

$$\begin{tikzcd} B \arrow[r, dash] \arrow[d, dash, "n^2"'] & V \\ F \arrow[ur, dash, "mn"'] \arrow[d, dash, "1 \text{ or } 2"'] & \\ F_0 \arrow[d, dash, "g"'] & \\ k & \end{tikzcd} \qquad \begin{tikzcd} & G' \arrow[d] \arrow[dr] & \\ Z(G) \arrow[r] \arrow[dr] & G_1 \arrow[r, "\mathrm{ad}"] \arrow[d, "\nu"] & G_1^{\mathrm{ad}} \\ & T_1 & \end{tikzcd} \qquad \begin{tikzcd} & G' \arrow[d] \arrow[dr] & \\ Z(G) \arrow[r] \arrow[dr] & G \arrow[r, "\mathrm{ad}"] \arrow[d, "\nu"] & G^{\mathrm{ad}} \\ & T & \end{tikzcd}$$

We let $*$ denote the adjoint involution on $\mathrm{End}_k(V)$ with respect to ψ — by assumption, it induces $*$ on B. Let C denote the commutant $\mathrm{End}_B(V)$ of B in $\mathrm{End}_F(V)$. Then C is a semisimple algebra, and it is stable $*$ on $\mathrm{End}_k(V)$. The groups G_1, G, and G' have the following descriptions:

$$\begin{aligned} G_1(k) &= \{x \in C \mid x^*x \in F_0^{\times}\} \\ G(k) &= \{x \in C \mid x^*x \in k^{\times}\} \\ G'(k) &= \{x \in C \mid x^*x = 1,\ \mathrm{Nrd}(x) = 1\}. \end{aligned}$$

EXAMPLE 8.5. Let F be an étale k-algebra of degree 2. Thus F is $k \times k$ or a field of degree 2 over k, and there is a unique nontrivial involution $*$ of F fixing the elements of k. We determine the symplectic $(B,*)$-modules in the case that B is isomorphic to a matrix algebra over F and $(B,*)$ is of type (A). These conditions imply that $B \simeq \mathrm{End}_F(W)$, where W is any simple B-module, and $*$ is the adjoint involution of a hermitian form $\phi{:}\, W \times W \to F$.

Let V_0 be a free F-module of finite rank, and let $\psi_0{:}\, V_0 \times V_0 \to F$ be a skew-hermitian form. Then B acts on $V \overset{\text{def}}{=} W \otimes_F V_0$ through the first factor, and we let ψ denote the k-bilinear form $V \times V \to k$ such that

$$\psi(w \otimes v, w' \otimes v') = \mathrm{Tr}_{F/k}(\phi(w,w')\psi_0(v,v')). \tag{46}$$

Then (V,ψ) is a symplectic $(B,*)$-module.

Conversely, let (V,ψ) be a symplectic $(B,*)$-module. As a B-module, V is a direct sum of copies of W, and so $V = W \otimes_F V_0$, where V_0 is a free F-module of finite rank. Let

$f \in F^{\times}$ be such that $f^* = -f$. According to Lemma A.7, there is a unique hermitian form $\Psi: V \times V \to F$ such that

$$\psi(u,v) = \mathrm{Tr}_{F/k}(f\Psi(u,v))$$

for all $u,v \in V$; moreover $\Psi(b^*u,v) = \Psi(u,bv)$ for all $b \in B$. The adjoint involution of Ψ on V preserves $\mathrm{End}_B(V) \simeq \mathrm{End}_F(V_0)$. We choose a skew-hermitian form $\psi_0: V_0 \times V_0 \to F$ whose adjoint involution is the restriction of $*_\Psi$ to $\mathrm{End}_F(V_0)$. The form $(u,v) \mapsto f\Psi(u,v)$ is skew-hermitian, and $f\Psi = \phi \otimes \psi_0$ (after possibly scaling ψ_0). Now ψ, ϕ, ψ_0 are related by (46).

We now determine the corresponding algebraic groups. We have

$$C \overset{\text{def}}{=} \mathrm{End}_B(V) \simeq \mathrm{End}_F(V_0)$$

and $*$ acts on C as the adjoint involution of the skew-hermitian form ψ_0. Therefore, G_1 (resp. G) is the group of unitary similitudes of ψ_0 whose multiplier lies in $F^{\times}$ (resp. $k^{\times}$), and G' is the special unitary group of ψ_0 (or the hermitian form $f\psi_0$).

EXAMPLE 8.6. We determine the symplectic $(B,*)$-modules in the case that B is isomorphic to a matrix algebra over k and $(B,*)$ is of type (C). These conditions imply that $B \simeq \mathrm{End}_k(W)$, where W is any simple B-module, and $*$ is the adjoint involution of a symmetric bilinear form $\phi: W \times W \to k$. Let V_0 be a k-vector space, and let ψ_0 be a skew-symmetric form $V_0 \times V_0 \to k$. Let ψ denote the k-bilinear form on $V \overset{\text{def}}{=} W \otimes V_0$ such that

$$\psi(w \otimes v, w' \otimes v') = \phi(w,w')\psi_0(v,v').$$

Then (V,ψ) is a symplectic $(B,*)$-module, and every such module arises in this way by an argument similar to that in 8.5. We have

$$C \overset{\text{def}}{=} \mathrm{End}_B(V) \simeq \mathrm{End}_k(V_0),$$

and $*$ acts on C as the adjoint involution of ψ_0. Therefore $G = \mathrm{GSp}(V_0,\psi_0)$ and $G' = \mathrm{Sp}(V_0,\psi_0)$.

PROPOSITION 8.7. *Let $(B,*)$ be a semisimple k-algebra with involution, and let (V,ψ) be a symplectic $(B,*)$-module. Let F denote the centre of B and F_0 its subalgebra of elements fixed by $*$. Assume that V and B are free over F. Set*

$$[B:F] = n^2, \quad [F_0:k] = g, \quad \dim_F(V) = mn.$$

We have, case-by-case.

(A) *The groups G and G_1 are connected and reductive, and G' is semisimple and simply connected; F is a quadratic extension of F_0, and*

$$\det(g)\cdot\det(g)^* = \mu(g)^m, \quad \textit{all } g \in G_1(k).$$

When k is algebraically closed, $G' \approx (\mathrm{SL}_m)^{[F_0:k]}$ (hence G' is of type A_{m-1}). If m is even, say $m = 2\ell$, then $(\det^{-1}\cdot\mu^{\ell},\mu)$ defines isomorphisms

$$T_1 \simeq \mathrm{Ker}\left((\mathbb{G}_m)_{F/k} \xrightarrow{\mathrm{Nm}} (\mathbb{G}_m)_{F_0/k}\right) \times (\mathbb{G}_m)_{F_0/k}$$

$$T \simeq \mathrm{Ker}\left((\mathbb{G}_m)_{F/k} \xrightarrow{\mathrm{Nm}} (\mathbb{G}_m)_{F_0/k}\right) \times \mathbb{G}_m.$$

If m is odd, say $m = 2\ell - 1$, then $\kappa \overset{\text{def}}{=} \det^{-1} \cdot \mu^{\ell}$ defines an isomorphism

$$T_1 \simeq (\mathbb{G}_m)_{F/k},$$

and $\mu = \kappa \cdot \kappa^*$.

(C) *The groups G and G_1 are connected and reductive, and G' is semisimple and simply connected. The integer m is even, say $m = 2\ell$, and*

$$\det(g) = \mu(g)^{\ell}, \quad \text{all } g \in G_1(k).$$

The map μ identifies T_1 with $(\mathbb{G}_m)_{F/k}$ and T with $\mathbb{G}_m$. When k is algebraically closed, $G' \approx (\mathrm{Sp}_m)^{[F:k]}$ (hence G' is of type $C_{m/2}$).

PROOF. It suffices to prove this after extending the base field k.

In case (A), we may suppose that B is isomorphic to a matrix algebra over F. Then $B \simeq \mathrm{End}_F(W)$, $V = W \otimes_F V_0$, and $\psi = \mathrm{Tr}_{F/k}(\phi \cdot \psi_0)$, as in (8.5). The statement can now be proved directly.

In case (C), we may suppose that B is isomorphic to a product of matrix algebras over k. Then $B \simeq \prod_{i \in I} \mathrm{End}_k(W_i)$ with $*$ on $\mathrm{End}_k(W_i)$ being the adjoint involution of a symmetric form ϕ_i. Moreover, $V \simeq \prod_{i \in I} W_i \otimes_k V_i$ and $\psi = \prod_{i \in I} \phi_i \otimes \psi_i$ with ψ_i a skew-symmetric form on V_i (cf. 8.6). The statement can now be proved directly. □

REMARK 8.8. We divide the case (BD) into the case (B) (m odd) and the case (D) (m even). In case (B), the groups are not of interest because they are not part of a Shimura datum.

REMARK 8.9. In case (D), the groups G and G_1 have $2^{[F:k]}$ connected components, and their identity components are reductive. The subgroup G' is semisimple (in particular, connected) but not simply connected. When k is algebraically closed, $G' \approx (\mathrm{SO}_m)^{[F:k]}$.

NOTES. This section follows Deligne 1971c, §5.

Algebras with positive involution

Let $(C, *)$ be a semisimple $\mathbb{R}$-algebra with involution, and let V be a C-module. In the next proposition, by a ***hermitian form*** on V we mean a C-balanced symmetric $\mathbb{R}$-bilinear form $\psi\colon V \times V \to \mathbb{R}$. For example, if $C = \mathbb{C}$ and $*$ is complex conjugation, then such a form can be written uniquely as $\psi = \mathrm{Tr}_{\mathbb{C}/\mathbb{R}} \circ \phi$ with $\phi\colon V \times V \to \mathbb{C}$ a hermitian form in the usual sense (see (8), p. 8). Such a form ψ 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)⇒(b). Let V be a faithful C-module. Every simple C-module occurs as a direct summand of V, and so every C-module occurs as a direct summand of a direct sum of copies of V. Hence, if V carries a positive-definite hermitian form, then so does every C-module.

(b)⇒(c). Let V be a C-module with a positive-definite hermitian form $(|)$, and choose an orthonormal $\mathbb{R}$-basis $e_1,\ldots,e_n$ for V. Then

$$\mathrm{Tr}_{\mathbb{R}}(c^*c|V) = \sum\nolimits_i (e_i|c^*ce_i) = \sum\nolimits_i (ce_i|ce_i),$$

which is > 0 unless c acts as the zero map on V. On applying this remark with $V = C$, we obtain (c).

(c)⇒(a). Condition (c) says that the hermitian form $(c,c') \mapsto \mathrm{Tr}_{C/\mathbb{R}}(c^*c')$ on the (faithful) C-module 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 positive involution and let (V,ψ) a symplectic $(B,*)$-module. Assume that $(B,*)$ is of type (A) or (C) and let C be the centralizer of B in $\mathrm{End}_{\mathbb{R}}(V)$. Then there exists a homomorphism of $\mathbb{R}$-algebras $h{:}\,\mathbb{C} \to C$ 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

$$J^2 = -1, \quad \psi(Ju,Jv) = \psi(u,v), \quad \psi(v,Jv) > 0 \text{ if } v \neq 0, \tag{47}$$

i.e., a complex structure satisfying the period relations p. 72.

Suppose first that $(B,*)$ is of type (A). Then $(B,*,V,\psi)$ decomposes into a product of systems as in (8.5). Thus, we may suppose that

◇ $B = \mathrm{End}_{\mathbb{C}}(W)$,

◇ $*$ is the adjoint involution of a positive-definite hermitian form $\phi{:}\,W \times W \to \mathbb{C}$,

◇ $V = W \otimes_{\mathbb{C}} V_0$ with V_0 a $\mathbb{C}$-vector space, and

◇ $\psi = \mathrm{Tr}_{\mathbb{C}/\mathbb{R}}(\phi \cdot \psi_0)$ with ψ_0 a skew-hermitian form on V_0.

We then have to classify the $J \in C \simeq \mathrm{End}_{\mathbb{C}}(V_0)$ satisfying (47) with ψ replaced by ψ_0. There exists a basis (e_j) for V_0 such that

$$(\psi_0(e_j,e_k))_{j,k} = \mathrm{diag}(\underset{1}{i},\ldots,\underset{r}{i},-i,\ldots,-i), \quad i = \sqrt{-1}.$$

Define J by $J(e_j) = -\psi_0(e_j,e_j)e_j$, i.e.,

$$J(e_j) = \begin{cases} -ie_j & \text{if } j \leq r \\ ie_j & \text{if } j > r. \end{cases}$$

Then J satisfies (47) with ψ_0 for ψ. This proves the result for type (A).

The proof for $(B.*)$ is similar (but easier). □

Let $(B,*)$ be a semisimple $\mathbb{R}$-algebra with positive involution and let (V,ψ) a symplectic $(B,*)$-module. Let $h{:}\,\mathbb{S} \to G_{\mathbb{R}}$ be a homomorphism such that (V,h) is of type

$\{(-1,0),(0,-1)\}$ and $(u,v)\mapsto\psi(u,h(i)v)$ is symmetric and positive definite. There are canonical isomorphisms of complex vector spaces

$$(V,J)\simeq V(\mathbb{C})/F_h^0V(\mathbb{C})\simeq V^{-1,0},$$

where $J=h(i)$. These are compatible with the actions of B, and we define $t(b)$ to be the trace of $b\in B$ on any one of these spaces. Thus

$$t(b)=\mathrm{Tr}_{\mathbb{C}}(b|(V,J)),\quad b\in B,\quad J=h(i).$$

PROPOSITION 8.13 (DELIGNE 1971C, 5.10). *With the above notation,*

(a) *the $G'(\mathbb{R})$ conjugacy class of h is uniquely determined by the map $t\colon B\to\mathbb{C}$;*

(b) *in case (A), the isomorphism class of (V,ψ) is determined by t; in cases (C) and (D), it is determined by $\dim_k(V)$ alone.*

(c) *The centralizer of h in $G(\mathbb{R})$ and $G_1(\mathbb{R})$ is connected.*

PROOF. These can be proved case by case. 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). □

An involution $*$ of a semisimple algebra B over $\mathbb{Q}$ is said to be ***positive*** if $\mathrm{Tr}_{B/\mathbb{Q}}(b^*b)>0$ for all nonzero $b\in B$. This is equivalent to requiring that $*$ becomed positive on $B\otimes_{\mathbb{Q}}\mathbb{R}$.

PEL data

We fix a simple algebra B over $\mathbb{Q}$ with a positive involution $*$. The centre of B is a number field F, and we let F_0 denote the subfield of elements fixed by $*$. Let $[B\colon F]=n^2$ and $[F_0\colon k]=g$. We assume throughout that $(B,*)$ is either of type (A) or type (C). In the first case, the involution is of the second kind and the field F_0 is totally real (because the trivial involution is positive). In the second case, the involution is of the first kind, and for one (hence every) splitting field L of B and isomorphism $B\otimes_F L\to M_n(L)$, the involution $*$ on $B\otimes_F L$ corresponds to an involution $x\mapsto c^{-1}\cdot x^t\cdot c$ on $M_n(L)$ with $c^t=c$;[53] in this case, $F=F_0$ is totally real.

Let (V,ψ) be symplectic $(B,*)$-module, and let $\dim_F(V)=mn$. Define G_1, G, and G' as before.

PROPOSITION 8.14 (ZINK 1983, 3.1). *There exists a homomorphism $h\colon\mathbb{S}\to G_{\mathbb{R}}$ such that (V,h) has type $\{(-1.0),(0.-1)\}$and $2\pi i\psi$ is a polarization of (V,h); moreover, h is unique up to conjugation by an element of $G(\mathbb{R})$.*

PROOF. We consider the decomposition

$$V_{\mathbb{R}}=\bigoplus\nolimits_{\sigma\colon F_0\to\mathbb{R}}V\otimes_{F_0,\sigma}\mathbb{R}.$$

[53] In other words, $B\otimes_F K\approx\mathrm{End}_F(W)$ and $*$ corresponds to the adjoint involution of a symmetric bilinear form.

Obviously, the summands are pairwise orthogonal under $\psi = \bigoplus_{\sigma} \psi_{\sigma}$. As h commutes with the action of $F_0 \otimes \mathbb{R}$, we have $h = \bigoplus h_{\sigma}$. Therefore, we need only consider a direct summand $V \otimes_{F_0,\sigma} \mathbb{R}$. Our problem is then equivalent to finding a complex structure $J: V \otimes_{F_0,\sigma} \mathbb{R} \to V \otimes_{F_0,\sigma} \mathbb{R}$, $J^2 = -1$, that commutes with the action of $B \otimes_{F_0,\sigma} \mathbb{R}$, and satisfied the period relations

$$\psi_{\sigma}(Ju, Jv) = \psi_{\sigma}(u, v),$$
$$\psi_{\sigma}(v, Jv) > 0.$$

Such a J is constructed in 8.12.

For type (A), let r_{ρ} be the number of i and $r_{\bar{\rho}}$ the number of $-i$ in the normal form of ψ_0. Then

$$t(b) \stackrel{\text{def}}{=} \mathrm{Tr}_{\mathbb{C}}(x \mid V_{\mathbb{R}}) = \sum_{\rho: F \to \mathbb{C}} r_{\rho} \rho(\mathrm{Tr}(b)), \quad b \in B.$$

Here $V_{\mathbb{R}}$ is to be understood with the complex structure J and Tr is the reduced trace. We remark that, conversely, r_{ρ} determines the symplectic $B \otimes \mathbb{R}$-module $(V \otimes \mathbb{R}, \Psi)$ up to isomorphism and it determines h up to conjugation (8.13).

For type (C), we obtain for the trace

$$t(x) = (m/2)\,\mathrm{Tr}^0_{B/\mathbb{Q}}\, x. \qquad \square$$

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.

The trace function

$$t(b) = \mathrm{Tr}(b|V(\mathbb{C})/F^0_h(V(\mathbb{C})), \quad h \in X,$$

depends only on X and it determines X. The pair (G, X) satisfies the conditions SV1–4.

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 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

Let $(B, *)$ be a semisimple $\mathbb{Q}$-algebra with involution and (V, ψ) a faithful $(B, *)$-symplectic module. Let G be the algebraic subgroup of $\mathrm{GL}_B(V)$ such that

$$G(\mathbb{Q}) = \{g \in \mathrm{GL}_B(V) \mid \psi(gx, gy) = \mu(g) \cdot \psi(x, y) \text{ for some } \mu(g) \in \mathbb{Q}^{\times}\}$$

The identity component of G is a reductive group, but G is not necessarily connected (see 8.7, 8.8, 8.9). Assume that there exists a homomorphism $h: \mathbb{S} \to G_{\mathbb{R}}$ such that (V, h) has type $\{(-1,0), (0,-1)\}$ and the form $\psi(u, h(i)v)$ is symmetric and positive-definite, and let X denote its $G(\mathbb{R})$-conjugacy class. The action of G on V defines a homomorphism $G \hookrightarrow G(\psi)$ which sends X into $X(\psi)$, and so (G,X) satisfies the conditions SV1–4. When G is connected,[54] we call (G,X) a ***PEL Shimura datum.***

THEOREM 8.17. *Let (G,X) be PEL Shimura datum, as above, 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 $((A,i),s,\eta K)$, where*

- ◇ *A is a complex abelian variety,*
- ◇ *$\pm s$ is a polarization of the Hodge structure $H_1(A,\mathbb{Q})$,*
- ◇ *i is a homomorphism $B \to \mathrm{End}^0(A)$, and*
- ◇ *ηK is a K-orbit of $B \otimes \mathbb{A}_f$-linear isomorphisms $\eta\colon V(\mathbb{A}_f) \to V_f(A)$ 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, and for such an isomorphism $a \circ h_A \circ a^{-1} \in X$.*

PROOF. In view of the dictionary $b \leftrightarrow t_b$ between endomorphisms and tensors (8.16), this follows from Theorem 7.4 □

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

$$\mathrm{Tgt}_0(A) \overset{(41)}{\simeq} H_1(A,\mathbb{C})/F^0 \xrightarrow{a} V(\mathbb{C})/F_h^0.$$

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 $B \otimes_{\mathbb{Q}} \mathbb{Q}^{\mathrm{a}}$-isomorphism $\alpha\colon V(\mathbb{Q}^{\mathrm{a}}) \to W(\mathbb{Q}^{\mathrm{a}})$ sending t to a $\mathbb{Q}^{\mathrm{al}\times}$-multiple of ψ. For $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ write $\sigma\alpha = \alpha \circ a_\sigma$ with $a_\sigma \in G(\mathbb{Q}^{\mathrm{a}})$. 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 $\alpha \circ a^{-1}$

[54] This hypothesis is only necessary so that $\mathrm{Sh}(G,X)$ is a Shimura variety in the sense of Deligne 1979. As noted elsewhere, much of the theory goes through for nonconnected G. Of course, one can replace G with its identity component, but then the theorem fails.

is fixed by all $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$, 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.

Finally, the fact that every isomorphism a makes h_A correspond to an element of X follows from 8.13(a). □

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

$$\begin{array}{ccccccccc}
 & & T(\mathbb{Q}) & \longrightarrow & H^1(\mathbb{Q}, G^{\mathrm{der}}) & \longrightarrow & H^1(\mathbb{Q}, G) & \longrightarrow & H^1(\mathbb{Q}, T) \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
G(\mathbb{R}) & \longrightarrow & T(\mathbb{R}) & \longrightarrow & H^1(\mathbb{R}, G^{\mathrm{der}}) & \longrightarrow & \prod_l H^1(\mathbb{Q}_l, G) & \longrightarrow & \prod_l H^1(\mathbb{Q}_l, T).
\end{array}$$

If an element c of $H^1(\mathbb{Q}, G)$ becomes trivial in $H^1(\mathbb{Q}_l, G)$ for all l, 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),

$$T = \mathrm{Ker}\left((\mathbb{G}_m)_F \xrightarrow{\mathrm{Nm}_{F/k}} (\mathbb{G}_m)_{F_0} \right) \times \mathbb{G}_m.$$

The group $H^1(\mathbb{Q}, \mathbb{G}_m) = 0$, and the map on H^1's of the first factor is

$$F_0^\times / \mathrm{Nm}\, F^\times \to \prod\nolimits_v F_{0v}^\times / \mathrm{Nm}\, F_v^\times.$$

This is injective because F/F_0 is cyclic (CFT, VIII 3.1).

For G of type (C), $T = \mathbb{G}_m$, and so $H^1(\mathbb{Q}, T) = 0$. □

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 1971c, §§5,6. See also Zink 1983, Langlands and Rapoport 1987, §6, and Kottwitz 1992, §§1–4.

9 General Shimura varieties

By definition, the Shimura varieties of Hodge type are those that are moduli varieties for abelian varieties with tensor and level structures. In order to realize more Shimura varieties as moduli varieties, we must enlarge the category of objects considered. We define a category of abelian motives that is generated by abelian varieties, and we define the class of Shimura varieties said to be of abelian type. Those Shimura varieties of abelian type with rational weight are moduli varieties for abelian motives. Finally, we note that the Shimura varieties of abelian type do not exhaust the 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. Moreover is semisimple because the polarization allows us to define a complement to a Hodge substructure.

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,

$$H^r(M_1 \sqcup M_2, \mathbb{Q}) \simeq H^r(M_1, \mathbb{Q}) \oplus H^r(M_2, \mathbb{Q}).$$

For each connected component V° of V,

$$H^*(V^\circ, \mathbb{Q}) \simeq \bigwedge H^1(V^\circ, \mathbb{Q}) \simeq \mathrm{Hom}_{\mathbb{Q}}(\bigwedge H_1(V^\circ, \mathbb{Q}), \mathbb{Q})$$

(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

$$\begin{aligned}
(V, e, m) \oplus (V', e', m) &= (V \sqcup V', e \oplus e', m) \\
(V, e, m) \otimes (V', e', m) &= (V \times V', e \otimes e', m + m') \\
(V, e, m)^\vee &= (V, e^t, d - m) \text{ if } V \text{ is purely } d\text{-dimensional.}
\end{aligned}$$

Here e^t denotes the transpose of e regarded as a correspondence.

For an abelian motive (V, e, m) over $\mathbb{C}$, let $H(V, e, m) = eH^*(V, \mathbb{Q})(m)$. Then

$$(V, e, m) \rightsquigarrow H(V, e, m)$$

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. Note that $\mathbb{Q}(1) \simeq \bigwedge^2 H_1(A,\mathbb{Q})$ for A an elliptic curve, and so it is abelian. We let $\mathsf{Hod}^{\mathrm{ab}}(\mathbb{Q})$ denote the full subcategory of $\mathsf{Hod}(\mathbb{Q})$ of abelian Hodge structures.

PROPOSITION 9.1. *The category $\mathsf{Hod}^{\mathrm{ab}}(\mathbb{Q})$ 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. □

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 J on $V(\mathbb{R})$ such that $\psi(Ju,Jv) = \psi(u,v)$ and $\psi(u,Ju) > 0$.

DEFINITION 9.2. (a) A connected Shimura datum (H,X^+) is of ***primitive abelian type*** if H is simple and there exists a symplectic space (V,ψ) over $\mathbb{Q}$ 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 connected Shimura data (H_i,X_i^+) of primitive abelian type and an isogeny $\prod_i H_i \to H$ carrying $\prod_i X_i^+$ into X^+.

(c) A Shimura datum (G,X) is of ***abelian type*** if (G^{der},X^+) is of abelian type.

(d) A Shimura variety $\mathrm{Sh}(G,X)$ is of ***abelian type*** if (G,X) is of abelian type (and similarly for connected Shimura varieties).[55]

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 $(V,\rho\circ h)$ 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 $(V,\rho\circ h)$ is an abelian Hodge structure for all h, then (G,X) is of abelian type.

PROOF. See Milne 1994b, 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 $g \in G$ and a fixed m, and

(b) ψ is a polarization of $(V,\rho\circ h)$ for all $h \in X$.

[55]The Shimura varieties of abelian type are exactly those proved to have canonical models in Deligne 1979. The name, which was introduced in Milne and Shih 1979, was chosen because of the relation of the varieties to abelian varieties. Only later was it realized that those with rational weight are the moduli of abelian motives (Milne 1994b).

Then there exist multilinear maps $t_i\colon V \times \cdots \times V \to \mathbb{Q}(r_i)$, $1 \leq i \leq n$, such that G is the subgroup of $\mathrm{GL}(V)$ whose elements satisfy (a) and fix $t_1, \ldots t_n$ (cf. 7.2).

THEOREM 9.4. *With the above notation,* $\mathrm{Sh}(G,X)$ *classifies the isomorphism classes of triples* $(A,(s_i)_{0\leq i\leq n},\eta K)$ *in which*

- ◇ A *is an abelian motive,*
- ◇ $\pm s_0$ *is a polarization for the rational Hodge structure* $H(A)$,
- ◇ $s_1,\ldots,s_n$ *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 Proposition 9.3 shows that the Hodge structures arising are abelian, and so they can be replaced by abelian motives (9.1). For more details, see Milne 1994b, Theorem 3.31. □

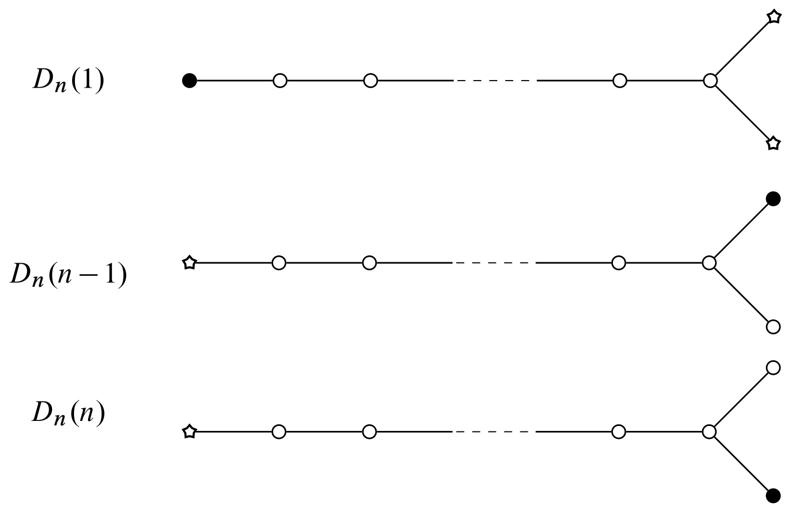
Classification of Shimura varieties of abelian type

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 because there is no symplectic embedding. If G^{ad} is of type D, (G,X^+) may or may not be of abelian type. In this last case, 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:



Solid nodes are special (p. 19), 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 solid node. Since X^+ is defined over $\mathbb{R}$, the solid 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 star. Note that the star has to be at the opposite end of the diagram from the solid node. In order for a family of representations $G_{i\mathbb{R}} \to S(\psi)_{\mathbb{R}}$, $1 \le i \le r$, to arise from a symplectic representation over $\mathbb{Q}$, the stars 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 solid nodes occur at different ends. See Deligne 1979, 2.3, or Milne 2013, §10, for more details.

In Addendum B, we provide a list of Shimura data (G, X) such that every connected Shimura datum of primitive abelian type is of the form (G^{der}, X^+) for some pair on the list.

Shimura varieties not of abelian type

It is hoped (Deligne 1979, p. 248)[56] 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 $\rho_{\mathbb{R}} \circ h$ 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 cannot.

[56]Pour interpréter des structures de Hodge de type plus compliqué, on aimerait remplacer les variétés abéliennes par de "motifs" convenables, mais il ne s'agit encore que d'un rêve. (To interpret Hodge structures of a more complicated type, one would like to replace the abelian varieties with suitable "motives", but this remains only a dream.)

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 Example 5.24. With the notation of that example,

$$G(\mathbb{R}) \approx \prod_{v\in I_c} \mathbb{H}^\times \times \prod_{v\in I_{nc}} \mathrm{GL}_2(\mathbb{R}).$$

(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 $d = [F:\mathbb{Q}]$ and i is a homomorphism $i: F \to \mathrm{End}(A)\otimes\mathbb{Q}$. These Shimura varieties are called ***Hilbert*** (or ***Hilbert-Blumenthal***) varieties. They form a natural first generalization of elliptic modular curves, and there are several books on them and their associated modular forms.[57]

(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 $2[F:\mathbb{Q}]$ and i is a homomorphism $i: B \to \mathrm{End}(A)\otimes\mathbb{Q}$. 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 $(F\otimes\mathbb{R})^\times \simeq \prod_{v:F\to\mathbb{R}}\mathbb{R}$ 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 $w_X:\mathbb{G}_m \to T_\mathbb{R}$ is defined over the subfield L of $\bar{\mathbb{Q}}$ whose fixed group is the subgroup of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ 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).

(d) When $|I_{nc}| = 1$, the Shimura variety is a curve. These are the famous Shimura curves.

ASIDE. A Shimura variety Sh_K that is a moduli variety will be a *coarse* moduli variety if K is insufficiently small or the condition SV5 fails (i.e., $Z(\mathbb{Q})$ is not discrete in $Z(\mathbb{A}_f)$) because then the objects classified may admit automorphisms. For example, an abelian variety, even a polarized abelian variety, may admit nontrivial automorphisms unless endowed with a level structure of level at least 3. The condition SV5 fails for Hilbert modular varieties because a totally real field of degree at least 2 has infinitely many units congruent to 1 modulo N for every N. In these situations, one may consider the ***Shimura stack*** instead. Form the complex analytic stack

$$\mathrm{Sh}_K^{\mathrm{an}}[G,X] = \left[G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K\right].$$

The groupoid of its complex points is equivalent to the following groupoid:

$$\begin{cases} \text{objects:} & \text{pairs } (x,g) \text{ in } X\times G(\mathbb{A}_f); \\ \text{morphisms } (x_0,g_0)\to(x_1,g_1): & \text{pairs } (q,k) \text{ in } G(\mathbb{Q})\times K \text{ with } qx_0 = x_1 \text{ and } qg_0k = g_1. \end{cases}$$

When the condition SV5 holds, the stabilizers in $\mathrm{Sh}_K^{\mathrm{an}}[G,X]$ are finite, and this complex analytic stack can be given the structure of a smooth Deligne–Mumford stack over $\mathbb{C}$. To see this, let K_0 be an open normal subgroup of K such that $K_0\cap G(\mathbb{Q})$ is neat (cf. 3.5). The finite group K/K_0 acts on $\mathrm{Sh}_K(G,X)$, and we can define

$$\mathrm{Sh}_K[G,X] = [\mathrm{Sh}_{K_0}(G,X)/(K/K_0)].$$

See Taelman 2017.

[57] Freitag, Hilbert modular forms, 1990; van der Geer, Hilbert modular surfaces, 1988; Garrett, Holomorphic Hilbert modular forms, 1990; Goren, Lectures on Hilbert modular varieties and modular forms, 2002.

Shimura varieties as moduli varieties[58]

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.

Summary

As we noted in §8, in general a (naturally-defined) moduli variety for abelian varieties is not a Shimura variety but only a finite union of Shimura varieties because the group G is not connected. Probably, the definition of a Shimura variety should be relaxed to allow G to be a nonconnected reductive group.[59] Also, the study of the boundaries of Shimura varieties suggests that the definition of a Shimura datum should be relaxed to allow X to be a finite covering of a conjugacy class of homomorphisms $\mathbb{S}\to G_{\mathbb{R}}$. Then a "Shimura variety of dimension zero" will indeed be a Shimura variety.

Shimura varieties

SVs of abelian type   SVs with rational weight

SVs of abelian type with rational weight (moduli varieties)

SVs of Hodge type

SVs of PEL type   Moduli varieties of PEL type

SVs of simple PEL type A, C (moduli varieties)

[58]I should include much more on this topic. See Milne 2013.

[59]In his Bourbaki talk (1971b) Deligne allowed a reductive group to be nonconnected, but in his Corvallis article (1979) he requires G to be connected, and it is the second article that became the fundamental reference.

10 Complex multiplication: the Shimura–Taniyama formula

Where we are headed

Let V be a variety over $\mathbb{Q}$. For any $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ and $P \in V(\mathbb{Q}^{\mathrm{a}})$, the point σP lies in $V(\mathbb{Q}^{\mathrm{a}})$. 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

$$f(a_1,\ldots,a_n)=0 \quad (a_i \in \mathbb{Q}^{\mathrm{a}}) \implies f(\sigma a_1,\ldots,\sigma a_n)=0$$

(apply σ to the first equality). Therefore, if we have a variety V over $\mathbb{Q}^{\mathrm{a}}$ that we suspect is actually defined over $\mathbb{Q}$, then we should be able to describe an action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ on its points $V(\mathbb{Q}^{\mathrm{a}})$.

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 canonical way) over specific number fields, and so I should be able to describe an action on their points of a subgroup of finite index in $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$. Suppose, for example, that our Shimura variety is of Hodge type. For each K, there is a set $\mathcal{M}_K$ whose elements are abelian varieties plus additional data and a map

$$(A,\ldots) \mapsto P(A,\ldots)\colon \mathcal{M}_K \to \mathrm{Sh}_K(\mathbb{C})$$

whose fibres are the isomorphism classes in $\mathcal{M}_K$ (7.4). 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 number field E, this does give an action of $\mathrm{Aut}(\mathbb{C}/E)$ on $\mathrm{Sh}_K(G,X)$ and that the action does indeed arise from a model of $\mathrm{Sh}_K(G,X)$ over E.

For example, we have surjective map

$$\{\text{elliptic curves over } \mathbb{C}\} \to \Gamma(1)\backslash\mathcal{H}_1$$

whose fibres are the isomorphism classes of elliptic curves (send $\mathbb{C}/(\mathbb{Z}+\mathbb{Z}\tau)$ to $[\tau]$). How does the natural action of $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ on the left transfer to an action on the right. If A maps to P, then I claim that σA maps to the unique point ${}^{\sigma}P$ such that $j({}^{\sigma}P)=\sigma(j(P))$. Here j is the usual j-function on $\mathcal{H}_1$. Indeed, if A is the curve $Y^2=X^3+aX+b$, then

$$j(A) \overset{\text{def}}{=} \frac{1728(4a^3)}{4a^3+27b^2} \text{ and } j(\sigma A)=\frac{1728(4\sigma a^3)}{4\sigma a^3+27\sigma b^2}=\sigma j(A).$$

But $j(A)=j(P)$ and $j(\sigma A)=j({}^{\sigma}P)$, which proves the claim. 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, and ${}^{\sigma}P=\sigma P$, but this is definitely false (if σ is not complex conjugation, then it is not continuous, nor even measurable). Unless A has complex multiplication, the only way to describe ${}^{\sigma}P$ is as the point such that $j({}^{\sigma}P)=\sigma(j(A))$ for every elliptic curve A mapping to P.

You may complain that the action of $\mathrm{Aut}(\mathbb{C}/E)$ on $\mathrm{Sh}_K(\mathbb{C})$ in the above examples is not explicit, but I contend that there cannot be a completely explicit description of the action. What are the elements of $\mathrm{Aut}(\mathbb{C}/E)$? To construct them, we have to choose a transcendence basis B for $\mathbb{C}$ over E, choose a permutation of the elements of B, and then choose an extension of the induced automorphism of $\mathbb{Q}(B)$ to its algebraic closure $\mathbb{C}$. These choices require the axiom of choice, 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 its 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 the maximal abelian extension of E in $\mathbb{C}$. If a point P has coordinates in E^{ab}, then the action of $\mathrm{Aut}(\mathbb{C}/E)$ on it factors through $\mathrm{Gal}(E^{\mathrm{ab}}/E)$, and so we may 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.

Briefly, when our Shimura variety arises naturally as a parameter variety (better, moduli variety) over $\mathbb{C}$, we can use the action on the $\mathbb{C}$-points to define a model of the variety over a specific number field. The theory of complex multiplication then gives us an explicit description of the action of $\mathrm{Aut}(\mathbb{C}/E)$ on certain special points (we call this the reciprocity law at the special point). The reciprocity laws at the special points determine the model uniquely. For a general Shimura variety, we define a model to be canonical if the "correct" reciprocity laws hold at the special points. We then use trickery to prove the existence of a canonical model.

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.

Let A be an abelian variety of dimension g over a field k. For all m not divisible by the characteristic of k,

$$A(k^{\mathrm{s}})_m \approx (\mathbb{Z}/m\mathbb{Z})^{2g}. \tag{48}$$

Here $A(k^{\mathrm{s}})_m$ is the set of elements of $A(k^{\mathrm{s}})$ killed by m. Hence, for $\ell \neq \mathrm{char}(k)$,

$$T_\ell A \stackrel{\text{def}}{=} \varprojlim A(k^{\mathrm{s}})_{\ell^n}$$

is a free $\mathbb{Z}_\ell$-module of rank $2g$, and

$$V_\ell(A) \stackrel{\text{def}}{=} T_\ell A \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$$

is a $\mathbb{Q}_\ell$-vector space of dimension $2g$. When k has characteristic zero, we set

$$T_f A = \prod\nolimits_\ell T_\ell A = \varprojlim_m A(k^{\mathrm{a}})_m,$$
$$V_f A = T_f \otimes_{\mathbb{Z}} \mathbb{Q} = \prod\nolimits_\ell (V_\ell A, T_\ell A) \text{ (restricted topological product).}$$

Thus $T_f A$ is a free $\hat{\mathbb{Z}}$-module of rank $2g$ and $V_f A$ is a free $\mathbb{A}_f$-module of rank $2g$. The Galois group $\mathrm{Gal}(k^{\mathrm{a}}/k)$ acts continuously and linearly on these modules.

For an endomorphism a of A, there is a unique monic polynomial $P_a(T)$ with integer coefficients (the ***characteristic polynomial of*** a) such that

$$|P_a(n)| = \deg(n - a)$$

for all $n \in \mathbb{Z}$. Moreover, P_a is the characteristic polynomial of a acting on $V_\ell A$ ($\ell \neq \mathrm{char}(k)$).

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 $A(\mathbb{C})_m \simeq \frac{1}{m}\Lambda/\Lambda \simeq \Lambda/m\Lambda$, and

$$T_\ell A \simeq \Lambda \otimes_{\mathbb{Z}} \mathbb{Z}_\ell, \quad V_\ell A \simeq \Lambda \otimes_{\mathbb{Z}} \mathbb{Q}_\ell, \quad T_f A = \Lambda \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}, \quad V_f A = \Lambda \otimes_{\mathbb{Z}} \mathbb{A}_f. \tag{49}$$

An endomorphism a of A defines a $\mathbb{C}$-linear endomorphism $(da)_0$ of $\mathrm{Tgt}_0(A)$ such that $(da)_0(\Lambda) \subset \Lambda$ (see 6.5), and $P_a(T)$ is the characteristic polynomial of $(da)_0$ on Λ.

For abelian varieties A and B over k, $\mathrm{Hom}(A,B)$ 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.[60] 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 is nonzero and 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:

$$\begin{array}{ccccc}
E \otimes_{\mathbb{Q}} \mathbb{R} & \xrightarrow{\simeq} & \prod_{\sigma:F\to\mathbb{R}} \mathbb{C}_\sigma & \qquad & \mathbb{C}_\sigma = E \otimes_{F,\sigma} \mathbb{R} \approx \mathbb{C} \\
\uparrow & & \uparrow & & \uparrow \\
a \otimes r \mapsto (\sigma a \cdot r)_\sigma \colon F \otimes_{\mathbb{Q}} \mathbb{R} & \xrightarrow{\simeq} & \prod_{\sigma:F\to\mathbb{R}} \mathbb{R}_\sigma & & \mathbb{R}_\sigma = \mathbb{R}
\end{array} \tag{50}$$

The involution $*$ is positive (in the sense of Definition 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

[60] Thus the objects of the skeleton of AV^0 are the isogeny classes of abelian varieties. Sometimes the objects of AV^0 itself are called "abelian varieties up to isogeny".

element from each conjugate pair $\{\varphi, \bar{\varphi}\}$. In other words, it is a subset $\Phi \subset \mathrm{Hom}(E, \mathbb{C})$ such that

$$\mathrm{Hom}(E,\mathbb{C}) = \Phi \sqcup \bar{\Phi} \qquad \text{(disjoint union, } \bar{\Phi} = \{\bar{\varphi} \mid \varphi \in \Phi\}).$$

Because E is quadratic over F, we have $E = F[\alpha]$ 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*** — note that its image in $\mathbb{C}$ under every embedding $E \hookrightarrow \mathbb{C}$ is purely imaginary.

The simplest CM-field is an imaginary quadratic extension E of $\mathbb{Q}$. A CM-type on E is the choice of an embedding $E \hookrightarrow \mathbb{C}$.

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

$$\mathrm{Tgt}_0(A) \approx \mathbb{C}^\Phi \quad \text{(as an } E \otimes_{\mathbb{Q}} \mathbb{C}\text{-module)}$$

for some CM -type Φ, then we say that (A,i) is of ***CM-type*** (E,Φ). Equivalently, (A,i) is of CM-type Φ if

$$\mathrm{Tr}(i(a) \mid \mathrm{Tgt}_0(A)) = \sum\nolimits_{\varphi \in \Phi} \varphi(a), \quad \text{all } a \in E. \tag{51}$$

REMARK 10.1. (a) In fact, (A,i) will always be of CM-type for some Φ. Recall (p. 73) that $A(\mathbb{C}) \simeq \mathrm{Tgt}_0(A)/\Lambda$ with Λ a lattice in $\mathrm{Tgt}_0(A)$. Then

$$\Lambda \otimes \mathbb{R} \simeq \mathrm{Tgt}_0(A),\ \Lambda \otimes \mathbb{Q} \simeq H_1(A,\mathbb{Q}),\ \Lambda \otimes \mathbb{R} \simeq H_1(A,\mathbb{R});$$
$$\Lambda \otimes \mathbb{C} \simeq H_1(A,\mathbb{C}) \simeq H^{-1,0} \oplus H^{0,-1} \simeq \mathrm{Tgt}_0(A) \oplus \overline{\mathrm{Tgt}_0(A)}.$$

Now $H_1(A,\mathbb{Q})$ is a one-dimensional vector space over E, and so $H_1(A,\mathbb{C}) \simeq \bigoplus_{\varphi: E \to \mathbb{C}} \mathbb{C}_\varphi$, where $\mathbb{C}_\varphi$ denotes a 1-dimensional vector space with E acting through φ. If φ occurs in $\mathrm{Tgt}_0(A)$, then $\bar{\varphi}$ occurs in $\overline{\mathrm{Tgt}_0(A)}$, and so $\mathrm{Tgt}_0(A) \simeq \bigoplus_{\varphi \in \Phi} \mathbb{C}_\varphi$ 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 is CM if A is simple, but not necessarily otherwise. For example, if B is an elliptic curve with complex multiplication by a quadratic field E, then every quadratic extension E' of E embeds into $M_2(E) \simeq \mathrm{End}^0(B \times B)$, but such an E' need not be a CM field.

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*

$$A_\Phi = \mathbb{C}^\Phi / \Phi(\mathcal{O}_E)$$

is an abelian variety 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

$$\mathcal{O}_E \otimes_{\mathbb{Z}} \mathbb{R} \simeq \mathcal{O}_E \otimes_{\mathbb{Z}} \mathbb{Q} \otimes_{\mathbb{Q}} \mathbb{R} \simeq E \otimes_{\mathbb{Q}} \mathbb{R} \xrightarrow[\simeq]{e \otimes r \mapsto (\ldots, \varphi e \cdot r, \ldots)} \mathbb{C}^\Phi,$$

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 (see 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 a positive integer 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 conditions for ψ to be a Riemann form can be checked on the right of (50). Here ψ takes the form $\psi = \sum_{\varphi\in\Phi} \psi_\varphi$, where

$$\psi_\varphi(u,v) = \mathrm{Tr}_{\mathbb{C}/\mathbb{R}}(\alpha_\varphi \cdot u \cdot \bar{v}), \quad \alpha_\varphi = \varphi(\alpha), \quad u,v \in \mathbb{C}.$$

Because α is totally imaginary,

$$\psi_\varphi(u,v) = \alpha_\varphi(u\bar{v} - \bar{u}v) \in \mathbb{R},$$

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 multiplication 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). This means that there exists an isomorphism $\mathbb{C}^\Phi \to \mathrm{Tgt}_0(A)$ of $E\otimes_{\mathbb{Q}}\mathbb{C}$-modules, and so $A(\mathbb{C})$ 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.). Therefore $\mathbb{Q}\Lambda = \Phi(E)\cdot\lambda$ for some $\lambda \in (E\otimes\mathbb{R})^\times$. After replacing the isomorphism $\mathbb{C}^\Phi \to \mathrm{Tgt}_0(A)$ by its composite with $\mathbb{C}^\Phi \xrightarrow{\lambda} \mathbb{C}^\Phi$, we may suppose 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 there are $\mathcal{O}_E$-isogenies

$$\mathbb{C}^\Phi/\Lambda \xrightarrow{N} \mathbb{C}^\Phi/N\Lambda \leftarrow \mathbb{C}^\Phi/\Phi(\mathcal{O}_E). \qquad \square$$

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).

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 $\mathbb{Q}^{\mathrm{a}}$, 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 (48)), 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 (see 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 $\mathbb{Q}^{\mathrm{a}}$. The polynomials defining A and i have coefficients in some subring R of $\mathbb{C}$ that is finitely generated over $\mathbb{Q}^{\mathrm{a}}$. According to the Hilbert Nullstellensatz, a maximal ideal $\mathfrak{m}$ of R will have residue field $\mathbb{Q}^{\mathrm{a}}$, and the reduction of (A,i) mod $\mathfrak{m}$ is called a ***specialization*** of (A,i). Every specialization (A',i') of (A,i) to a pair over $\mathbb{Q}^{\mathrm{a}}$ with A' nonsingular will still be of CM-type (E,Φ), because the CM-type is determined by the set of eigenvalues of a generator e of E over $\mathbb{Q}$ acting on the tangent space, and this set is unchanged by the change of ground ring from $\mathbb{C}$ to R to $\mathbb{Q}^{\mathrm{a}}$. Therefore, by Proposition 10.2, there exists an isogeny $(A',i')_{\mathbb{C}} \to (A,i)$. The kernel H of this isogeny is a subgroup of $A'(\mathbb{C})_{\mathrm{tors}} = A'(\mathbb{Q}^{\mathrm{a}})_{\mathrm{tors}}$, and $(A'/H,i)$ is a model of (A,i) over $\mathbb{Q}^{\mathrm{a}}$. □

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.[61]

Let A be an abelian variety over a number field K, and let $\mathcal{O}_{K,\mathfrak{P}}$ denote the localization of $\mathcal{O}_K$ at a prime ideal $\mathfrak{P}$. 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., to 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 the subring $\mathcal{O}_{K,\mathfrak{P}}[X_0,\ldots,X_n]$, and let $\bar{P}$ denote the reduction of P modulo $\mathfrak{P}$; the $\bar{P}$'s obtained in this fashion generate a homogeneous ideal $\bar{\mathfrak{a}}$ 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 $\bar{\mathfrak{a}}$ is an abelian variety $\bar{A}$ over k. Then $\bar{A}$ is called ***the reduction of*** A ***at*** $\mathfrak{P}$. It can be shown that, up to a canonical isomorphism, $\bar{A}$ is independent of all choices. For $\ell \neq \mathrm{char}(k)$, $V_\ell(A) \simeq V_\ell(\bar{A})$. There is an injective homomorphism $\mathrm{End}(A) \to \mathrm{End}(\bar{A})$ compatible with $V_\ell(A) \simeq V_\ell(\bar{A})$ (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 $\mathcal{O}_K$. After possibly replacing K by a finite extension, A will have good reduction at $\mathfrak{P}$.*

PROOF. We use the Néron–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 $V_\ell A \simeq H_1(A_{\mathbb{C}},\mathbb{Q}) \otimes \mathbb{Q}_\ell$ (see (49)). 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 $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ 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, 7.59), 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 the criterion applies. □

[61] Consider the curve

$$E : Y^2 + (j-1728)XY = X^3 - 36(j-1728)^2X - (j-1728)^3,$$

where $j \in \mathbb{C}$ is transcendental. Specializing E to $\mathbb{Q}^{\mathrm{a}}$ amounts to replacing j with an algebraic number, say, j', in the equation. Since E has j-invariant j, and the specialized curve E' has j-invariant j', we see that $E'_{\mathbb{C}}$ is not isomorphic to E.

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, for $a_1,\ldots,a_n \in \mathbb{F}$,

$$f(a_1,\ldots,a_n) = 0 \implies f(a_1^q,\ldots,a_n^q) = 0.$$

PROPOSITION 10.6. *There is a unique way of attaching to every variety V over $\mathbb{F}_q$ a regular map $\pi_V: V \to V$ such that*

(a) *for every regular map $\alpha: V \to W$, $\alpha \circ \pi_V = \pi_W \circ \alpha$;*
(b) *$\pi_{\mathbb{A}^1}$ is the map $a \mapsto a^q$.*

PROOF. For an affine variety $V = \operatorname{Spm} A$, 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 $\mathbb{F}_q$, $\operatorname{End}^0(A)$ 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. For a modern exposition of Weil's proof, see Milne 1986, 19.1. □

Moreover, $\operatorname{End}^0(A)$ is simple if and only if A is simple. Therefore, if A is simple, then $\mathbb{Q}[\pi_A]$ is a field, and π_A is an algebraic integer in it (p. 97). 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,

$$\rho(\pi)\cdot\overline{\rho(\pi)} = q = \rho(q) = \rho(\pi)\cdot\rho(q/\pi), \quad \text{all } \rho: \mathbb{Q}[\pi] \to \mathbb{C},$$

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 $\operatorname{ord}_v(\pi) = \operatorname{ord}_v(\pi')$ 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 $\operatorname{ord}_v(\pi) = 0$ 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 $(\bar{A},\bar{\imath})$ over $\mathcal{O}_k/\mathfrak{P}=\mathbb{F}_q$. Then the Frobenius map $\pi_{\bar{A}}$ of $\bar{A}$ lies in $\bar{\imath}(E)$.*

PROOF. Let $\pi=\pi_{\bar{A}}$. It suffices to check that π lies in $\bar{\imath}(E)$ after tensoring[62] 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 $V_\ell\bar{A}$ is also a free $E\otimes_{\mathbb{Q}}\mathbb{Q}_\ell$-module of rank 1 (via $\bar{\imath}$). Therefore, any endomorphism of $V_\ell\bar{A}$ 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,*

$$\frac{\mathrm{ord}_v(\pi)}{\mathrm{ord}_v(q)}=\frac{|\Phi\cap H_v|}{|H_v|},\tag{52}$$

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.

We sketch a proof of the theorem at the end of the section.

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.[63]

(b) Let $*$ denote complex conjugation on $\mathbb{Q}[\pi]$. Then $\pi\pi^*=q$, and so

$$\mathrm{ord}_v(\pi)+\mathrm{ord}_v(\pi^*)=\mathrm{ord}_v(q).\tag{53}$$

Moreover,

$$\mathrm{ord}_v(\pi^*)=\mathrm{ord}_{v^*}(\pi)$$

and

$$\Phi\cap H_{v^*}=\bar{\Phi}\cap H_v.$$

Therefore, (52) is consistent with (53):

$$\frac{\mathrm{ord}_v(\pi)}{\mathrm{ord}_v(q)}+\frac{\mathrm{ord}_v(\pi^*)}{\mathrm{ord}_v(q)}\overset{(52)}{=}\frac{|\Phi\cap H_v|+|\Phi\cap H_{v^*}|}{|H_v|}=\frac{|(\Phi\cup\bar{\Phi})\cap H_v|}{|H_v|}=1.$$

In fact, (52) is the only obvious formula for $\mathrm{ord}_v(\pi)$ consistent with (53), which is probably a more convincing argument for its validity than the proof sketched below.

[62] Let W be a subspace of a k-vector space V, and let R be a ring containing k. Then $(R\otimes_k W)\cap V=W$ (intersection inside V). To see this, note that an element v of V lies in W if and only if $f(v)=0$ for all $f\in(V/W)^\vee$, and that $f(v)$ is zero if and only if it is zero in R.

[63] Let π' arise from second model (A',i'). Then (A',i') will become E-isogenous to (A,i) over a finite extension k' of k (see 10.2), from which it follows that $\pi^f=\pi'^f$ for f the degree of the residue field extension.

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 $r_i \geq 1$, and the set (with multiplicities) of pairs $(\mathfrak{p}_i, r_i)$ is uniquely determined by M. Define[64] $|M|_R = \prod \mathfrak{p}_i^{r_i}$. 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, p. 68). 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,

$$|S/\mathfrak{A}|_R = \mathrm{Nm}(\mathfrak{A}) \tag{54}$$

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*

$$|\mathrm{Tgt}_0(A)|_{\mathcal{O}_E} = (\pi_A).$$

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 $\bar{A}$. 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

$$T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} k = \mathrm{Tgt}_0(A), \quad T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} \mathcal{O}_{k,\mathfrak{P}}/\mathfrak{P} = \mathrm{Tgt}_0(\bar{A}).$$

Therefore,

$$(\pi) \overset{10.12}{=} \left|\mathrm{Tgt}_0(\bar{A})\right|_{\mathcal{O}_E} = \left|T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} (\mathcal{O}_{k,\mathfrak{P}}/\mathfrak{P})\right|_{\mathcal{O}_E}. \tag{55}$$

For simplicity, assume[65] that $(p) \overset{\text{def}}{=} \mathfrak{P} \cap \mathbb{Z}$ is unramified in E. Then the isomorphism of E-modules

$$T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} k \to k^{\Phi}$$

restricts to an isomorphism of $\mathcal{O}_E$-modules[66]

$$T \to \mathcal{O}_{k,\mathfrak{P}}^{\Phi}. \tag{56}$$

---

[64] Better, the first statement shows that the K-group of the category of finitely generated torsion R-modules is canonically isomorphic to the group of fractional ideals of R, and so $|M|_R$ denotes the class of M in the K-group.

[65] This, in fact, is the only case we need, because it suffices for the proof of the main theorem in §10, which in turn implies the Shimura–Taniyama formula.

[66] Since $\mathcal{O}_E$ is unramified at p, $\mathcal{O}_E \otimes_{\mathbb{Z}} \mathbb{Z}_p$ is étale over $\mathbb{Z}_p$, and so $\mathcal{O}_E \otimes_{\mathbb{Z}} \mathcal{O}_{\mathfrak{P}}$ is étale over $\mathcal{O}_{\mathfrak{P}}$. In fact, the isomorphism $E \otimes_{\mathbb{Q}} k \simeq \prod_{\sigma: E \to k} k_\sigma$ induces an isomorphism $\mathcal{O}_E \otimes_{\mathbb{Z}} \mathcal{O}_{\mathfrak{P}} \simeq \prod_{\sigma: E \to k} \mathcal{O}_\sigma$ where $\mathcal{O}_\sigma$ denotes $\mathcal{O}_{\mathfrak{P}}$ regarded as a $\mathcal{O}_E$-algebra via σ. Thus, the finitely generated projective $\mathcal{O}_E \otimes_{\mathbb{Z}} \mathcal{O}_{\mathfrak{P}}$-modules are direct sums of $\mathcal{O}_\sigma$'s, from which the statement follows. For a more explicit proof, see Shimura 1999, 13.2.

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 φth copy through the map

$$\mathcal{O}_E \xrightarrow{\varphi} \mathcal{O}_k \subset \mathcal{O}_{k,\mathfrak{P}}.$$

As $\mathcal{O}_k/\mathfrak{P} \simeq \mathcal{O}_{k,\mathfrak{P}}/\mathfrak{P}$ (ANT, 3.10), the contribution of the φth copy to (π) in (55) is

$$|\mathcal{O}_k/\mathfrak{P}|_{\varphi\mathcal{O}_E} \overset{(54)}{=} \varphi^{-1}(\mathrm{Nm}_{k/\varphi E}\,\mathfrak{P}).$$

Thus,

$$(\pi) = \prod_{\varphi\in\Phi} \varphi^{-1}(\mathrm{Nm}_{k/\varphi E}\,\mathfrak{P}). \tag{57}$$

This is how Shimura and Taniyama state their formula (Shimura and Taniyama 1961, III, Theorem 1).

The derivation of (52) from (57) is easy. Because p is unramified in E, $\mathrm{ord}_v(p) = 1$ for all primes v of E dividing p, and so

$$\mathrm{ord}_v(q) = f(\mathfrak{P}/p).$$

The formula (57) can be restated as

$$\mathrm{ord}_v(\pi) = \sum_{\varphi\in\Phi\cap H_v} f(\mathfrak{P}/\varphi\mathfrak{p}_v),$$

and so

$$\frac{\mathrm{ord}_v(\pi)}{\mathrm{ord}_v(q)} = \sum_{\varphi\in\Phi\cap H_v} \frac{1}{f(\mathfrak{p}_v/p)} = |\Phi\cap H_v|\cdot\frac{1}{|H_v|}.$$

ASIDE. As Yoshida (BAMS 2002, p. 441) wrote, "Shimura's mathematics developed by stages:

(A) complex multiplication of abelian varieties $\Longrightarrow$

(B) the theory of canonical models = Shimura varieties $\Longrightarrow$

(C) critical values of zeta functions and periods of automorphic forms.

(B) includes (A) as the 0-dimensional special case of canonical models. The relation of (B) and (C) is more involved, but (B) provides a solid foundation of the notion of the arithmetic automorphic forms." In these notes, we treat only (A) and (B).

NOTES. The first statement of the Shimura-Taniyama formula that I know of is in Weil's conference talk (Weil 1956b, p. 21), where he writes "[For this] it is enough to determine the prime ideal decomposition of $\pi\ldots$ *But this has been done by Taniyama*" (italics in original). The formula (57) is proved in Shimura and Taniyama 1961, III.13, in the unramified case using spaces of differentials rather than tangent spaces. For a modern exposition of their proof, which is quite short and elementary, see II, 8, of my notes *Complex Multiplication*. These notes also discuss other proofs of the theorem (Giraud 1968, Tate 1968, Serre 1968).

11 Complex multiplication: the main theorem

Let E be an imaginary quadratic extension of $\mathbb{Q}$, and let A be an elliptic curve over $\mathbb{C}$ such that $\mathrm{End}(A) = \mathcal{O}_E$. Then $j(A)$ generates the Hilbert class field $\tilde{E}$ of E (maximum abelian unramified extension of E). Now let A be an elliptic curve over $\tilde{E}$ with $\mathrm{End}(A) = \mathcal{O}_E$, and suppose that $\mathcal{O}_E^\times = \{\pm 1\}$. Then the x-coordinates of the torsion points on A generate the maximum abelian extension E^{ab} of E. This is a brief statement of the classical theory of complex multiplication for elliptic curves (Kronecker, Weber, ...; see Serre 1967a). To prove these results one first determines how $\mathrm{Gal}(E^{\mathrm{ab}}/E)$ acts on A and its torsion points, and then one applies class field theory. Shimura and Taniyama (and Weil) extended the theory to abelian varieties. To a CM-field E and CM-type Φ on E, they attach a second CM-field E^*, called the reflex field. When E is quadratic over $\mathbb{Q}$, then $E^* = E$. Now let A be a complex abelian variety of CM-type (E, Φ). The main theorem of complex multiplication (11.2) describes how $\mathrm{Gal}(E^{*\mathrm{ab}}/E^*)$ acts on A and its torsion points. Later we shall re-interpret the theorem as describing how $\mathrm{Aut}(\mathbb{C}/E^*)$ acts on the complex points of the Shimura variety defined by a torus.

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^{a} 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

$$\begin{tikzcd} E^\times\backslash\mathbb{A}_E^\times \arrow[r, "\mathrm{rec}_E", "\text{onto}"'] \arrow[d] & \mathrm{Gal}(E^{\mathrm{ab}}/E) \arrow[d, "\sigma\mapsto\sigma|L"] \\ E^\times\backslash\mathbb{A}_E^\times/\mathrm{Nm}_{L/E}(\mathbb{A}_L^\times) \arrow[r, "\mathrm{rec}_{L/E}", "\simeq"'] & \mathrm{Gal}(L/E). \end{tikzcd}$$

It has the following properties (which determine it):

(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 the residue field $\mathcal{O}_L/\mathfrak{P}$.

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 the group $\mathrm{Gal}(E^{\mathrm{ab}}/E)$ 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

$$\begin{tikzcd}
\hat{\mathbb{Z}}^\times \arrow[r, "\mathrm{rec}_{\mathbb{Q}}", "\simeq"'] \arrow[d] & \mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}) \arrow[d, "\text{restrict}"] & \mathbb{Q}^{\mathrm{ab}} = \bigcup \mathbb{Q}[\zeta_N] \\
(\mathbb{Z}/N\mathbb{Z})^\times \arrow[r, "{[a] \mapsto (\zeta_N \mapsto \zeta_N^a)}", "\simeq"'] & \mathrm{Gal}(\mathbb{Q}[\zeta_N]/\mathbb{Q}) &
\end{tikzcd} \tag{58}$$

which *commutes with an inverse*, i.e., the two maps send an element of $\hat{\mathbb{Z}}^\times$ to inverse elements of $\mathrm{Gal}(\mathbb{Q}[\zeta_N]/\mathbb{Q})$. 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

$$\alpha = p \cdot (\underset{2}{1}, \ldots, 1, \underset{p}{p^{-1}}, 1, \ldots) \in \mathbb{Z} \cdot \mathbb{A}_f^\times = \mathbb{A}_f^\times.$$

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 *Class Field Theory*.

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

$$\mathrm{art}_E(\alpha) = \mathrm{rec}_E(\alpha)^{-1}. \tag{59}$$

Now, the diagram (58) commutes with $\mathrm{art}_{\mathbb{Q}}$ for $\mathrm{rec}_{\mathbb{Q}}$. 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 $\mathbb{Q}^{\mathrm{a}}$ characterized by any one of the following equivalent conditions:

(a) $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ 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 $\mathbb{Q}^{\mathrm{a}}$ such that there exists a k-vector space V with an action of E for which

$$\mathrm{Tr}_k(a|V) = \textstyle\sum_{\varphi\in\Phi} \varphi(a), \quad \text{all } a \in E.$$

ASIDE. We verify the equivalence.

(a)$\Leftrightarrow$(b). If $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ permutes the φ's, then clearly it fixes all elements of the form $\sum_{\varphi\in\Phi} \varphi(a)$. Conversely, if $\sum_{\varphi\in\Phi} \varphi(a) = \sum_{\varphi\in\Phi} (\sigma\varphi)(a)$ for all $a \in E^\times$, then $\{\sigma\varphi | \varphi \in \Phi\} = \Phi$ by Dedekind's theorem on the independence of characters. This shows that conditions (a) and (b) define the same field.

(b)$\Leftrightarrow$(c). If there exists a k-vector space V as in (c), then clearly k contains the field in (b). On the other hand, there exists a representation of $(\mathbb{G}_m)_{E/\mathbb{Q}}$ on a vector space V over the field E^* in (a) with Φ as its set of characters (p. 17), and the action of $E^\times = (\mathbb{G}_m)_{E/\mathbb{Q}}(\mathbb{Q})$ on V extends to an action of E with trace $\sum_{\varphi\in\Phi} \varphi(a)$.

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[67] $N_{\Phi^*}: (\mathbb{G}_m)_{E^*/\mathbb{Q}} \to (\mathbb{G}_m)_{E/\mathbb{Q}}$ 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 $\sigma: k \to \Omega$ 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 $i: E \to \mathrm{End}^0(A)$ defines a similar pair $\sigma(A,i) = (\sigma A, {}^\sigma i)$ over Ω. Here ${}^\sigma i: E \to \mathrm{End}(\sigma A)$ 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 $\sigma: V_f(A) \to V_f(\sigma A)$ provided that k and Ω are algebraically closed (otherwise one would have to choose an extension of k to a homomorphism $k^{\mathrm{a}} \to \Omega^{\mathrm{a}}$).

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 exists a unique E-linear 'isogeny' $\alpha: A \to \sigma A$ 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

$$\mathrm{Tgt}_0(\sigma A) = \mathrm{Tgt}_0(A) \otimes_{\mathbb{C},\sigma} \mathbb{C}$$

[67] One can show that E^* is again a CM-field, and that an embedding of E into $\mathbb{Q}^{\mathrm{a}}$ defines a CM-type on E^*. The reflex norm is usually defined in terms of Φ^* but we will not need it.

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' $\alpha: A \to \sigma A$ (see 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 $a \in \mathbb{A}_{E,f}^{\times}$. 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

$$\sigma \mapsto aE^{\times}: \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/E^*) \to \mathbb{A}_{E,f}^{\times}/E^{\times},$$

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

$$\eta: \mathbb{A}_{E^*,f}^{\times}/E^{*\times} \to \mathbb{A}_{E,f}^{\times}/E^{\times}.$$

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). Now $\alpha(N_{\Phi^*}(s)\cdot x) = \sigma x$ after possibly replacing α with a multiple of α by an element of $E^{\times}$. The uniqueness follows from the faithfulness of the functor $A \rightsquigarrow V_f(A)$. □

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 $\beta:(A,i) \to (B,j)$ 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, see §10 of Chapter II of my notes *Complex Multiplication*.

ASIDE. The Kronecker-Weber theorem says that $\mathbb{Q}^{\mathrm{ab}}$ can be obtained from $\mathbb{Q}$ by adjoining the special values $e^{2\pi i z}$, $z \in \mathbb{Q}^{\times}$, of the exponential function e. The classical theory of complex multiplication for elliptic curves says the maximum abelian extension of an imaginary quadratic field E can be obtained from E by adjoining certain special values of elliptic modular functions.

The twelfth of Hilbert's famous problems asked for an extension of these results to all number fields.

What answer does the theory of complex multiplication give to Hilbert's problem? This was worked out by Wafa Wei in her Ph.D. thesis (University of Michigan, 2003). Let E be a CM-field and let F be the totally real subfield of E such that $[E:F] = 2$. It is well-known that the theory of complex multiplication gives nothing about the abelian extensions of the totally real field F (except for those extensions coming from $\mathbb{Q}$). Subject to this limitation, it gives everything. The precise statement is the following. Let H be the image of the Verlagerung map $\mathrm{Gal}(F^{\mathrm{a}}/F) \to \mathrm{Gal}(E^{\mathrm{a}}/E)$. Then the extension of E obtained by adjoining the special values of all automorphic functions on the canonical models of all Shimura varieties with rational weight is $(E^{\mathrm{ab}})^H \cdot \mathbb{Q}^{\mathrm{ab}}$.

12 Definition of canonical models

Let (G, X) be a Shimura datum and (G', X^+) the associated connected Shimura datum. Attached to every compact open subgroup K of $G(\mathbb{A}_f)$, there is a map

$$f_K\colon \mathrm{Sh}_K(G, X) \to \pi_K$$

of varieties over $\mathbb{C}$. Here π_K is a variety of dimension zero, and the fibres of f_K are connected Shimura varieties attached to (G, X^+).

The theory of canonical models defines

(a) a "reflex" field $E = E(G, X)$, which is an algebraic number field contained in $\mathbb{C}$ depending only on (G, X);

(b) a "canonical" model $(f_K)_0\colon \mathrm{Sh}_K(G, X)_0 \to (\pi_K)_0$ of f_K over E, which is uniquely characterized by the reciprocity laws at the special points.

It also describes the action of $\mathrm{Aut}(\mathbb{C}/E)$ on π_K corresponding to its model $(\pi_K)_0$, and hence it defines a model of each connected component of $\mathrm{Sh}_K(G, X)$ over a finite extension of E.

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 $I_0 \overset{\text{def}}{=} I(V) \cap k[X_1, \ldots, X_n]$ 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

$$\mathbb{A}^2_{\mathbb{C}} \xleftarrow{(x,y) \leftarrow\!\shortmid (x,y)} V(X^2 + Y^2 - 1) \xrightarrow{(x,y) \mapsto (x, y/\sqrt{2})} \mathbb{A}^2_{\mathbb{C}}$$

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.

ASIDE. Most complex algebraic varieties have no model over a number field, so why do Shimura varieties have such a model? Here is a heuristic explanation. If the smallest field of definition of a complex algebraic variety is transcendental over $\mathbb{Q}$, then we can spread the variety out and obtain a flat family of varieties. Therefore, such a variety should not be locally rigid, i.e., it should admit nontrivial local deformations. This means that a complex algebraic variety should admit a model over $\mathbb{Q}^{\mathrm{a}}$ if it is locally rigid. The fact that there are only countably many arithmetic locally symmetric varieties up to isomorphism suggests that they are locally rigid and hence defined over $\mathbb{Q}^{\mathrm{a}}$. This heuristic argument can be made rigorous (Shimura, Faltings). For a recent exposition, see Peters 2017b,a.

The reflex field

The reflex field of a Shimura variety is a number field that is the "natural" field of definition of the Shimura variety. Indeed, whenever the Shimura is a moduli variety in some natural way, the reflex field is the field of definition of the moduli problem.

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:

$$\mathcal{C}(k) = G(k)\backslash \mathrm{Hom}(\mathbb{G}_m, G_k).$$

A homomorphism $k \to k'$ induces a map $\mathcal{C}(k) \to \mathcal{C}(k')$; in particular, $\mathrm{Aut}(k'/k)$ acts on $\mathcal{C}(k')$.

Assume that G splits over k, so that G_k contains a split maximal torus T. The Weyl group $W = W(G_k, T)$ is the constant étale algebraic group N/N°, where N is the normalizer of T in G_k. For every field k' containing k

$$W(k) = W(k') = N(k')/T(k')$$

(Milne 2017, 21.1).

LEMMA 12.1. *Let T be a split maximal torus in G_k. Then the map*

$$W\backslash \mathrm{Hom}(\mathbb{G}_m, T_k) \to G(k)\backslash \mathrm{Hom}(\mathbb{G}_m, G_k)$$

is bijective.

PROOF. As any two maximal split tori are conjugate (Milne 2017, 17.105), the map is surjective. Let μ and μ' be cocharacters of T that are conjugate by an element of $G(k)$, say, $\mu = \mathrm{ad}(g)\circ\mu'$ with $g \in G(k)$. Then $\mathrm{ad}(g)(T)$ and T are both maximal split tori in the centralizer[68] C of $\mu(\mathbb{G}_m)$, which is a connected reductive group (ibid., 17.59). Therefore, there exists a $c \in C(k)$ such that $\mathrm{ad}(cg)(T) = T$. Now cg normalizes T and $\mathrm{ad}(cg)\circ\mu' = \mu$, which proves that μ and μ' are in the same W-orbit. □

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 $\mathrm{Hom}(\mathbb{G}_m, T_{\mathbb{Q}^{\mathrm{a}}})$ nor W changes when we replace $\mathbb{C}$ with the algebraic closure $\mathbb{Q}^{\mathrm{a}}$ of $\mathbb{Q}$ in $\mathbb{C}$, and so the lemma shows that $c(X)$ contains a μ defined over $\mathbb{Q}^{\mathrm{a}}$ and that the $G(\mathbb{Q}^{\mathrm{a}})$-conjugacy class of μ is independent of the choice of μ. This allows us to regard $c(X)$ as an element of $\mathcal{C}(\mathbb{Q}^{\mathrm{a}})$.

[68]Certainly $T \subset C$. Let $t \in T(k^{\mathrm{a}})$ and $a \in \mathbb{G}_m(k^{\mathrm{a}})$. Then

$$gtg^{-1}\cdot\mu(a) = gt\cdot\mu'(a)\cdot g^{-1} = g\cdot\mu'(a)\cdot tg^{-1} = \mu(a)\cdot gtg^{-1},$$

and so $gTg^{-1} \subset C$. More conceptually, T centralizes both $\mu(\mathbb{G}_m)$ and $\mu'(\mathbb{G}_m)$, and the second condition implies that $\mathrm{ad}(g)(T)$ centralizes $\mathrm{ad}(g)\mu'(\mathbb{G}_m) = \mu(\mathbb{G}_m)$.

DEFINITION 12.2. The ***reflex*** (or ***dual) field*** $E(G,X)$ is the field of definition of $c(X)$ in $\mathbb{Q}^{\mathrm{a}}$, i.e., it is the fixed field of the subgroup of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ fixing $c(X)$ as an element of $\mathcal{C}(\mathbb{Q}^{\mathrm{a}})$ (or stabilizing $c(X)$ as a subset of $\mathrm{Hom}(\mathbb{G}_m, G_{\mathbb{Q}^{\mathrm{a}}})$).

We will see in 12.4(b) that this generalizes the reflex field of a CM type.

REMARK 12.3. (a) Any subfield k of $\mathbb{Q}^{\mathrm{a}}$ 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 $\mathbb{Q}^{\mathrm{a}}$. It 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. Then $i\colon G\to G'$ induces a $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$-equivariant map $\mathcal{C}_G(\mathbb{Q}^{\mathrm{a}})\to\mathcal{C}_{G'}(\mathbb{Q}^{\mathrm{a}})$ sending $c(X)$ to $c(X')$. Therefore,

$$E(G,X)\supset E(G',X').$$

EXAMPLE 12.4. We explain how to calculate the reflex field.

(a) Let T be a torus over $\mathbb{Q}$, and let h be a homomorphism $\mathbb{S}\to T_{\mathbb{R}}$. Then $\mu_h\colon\mathbb{G}_{m\mathbb{C}}\to T_{\mathbb{C}}$ is defined over $\mathbb{Q}^{\mathrm{a}}$, and $E(T,\{h\})$ is the fixed field of the subgroup of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ fixing $\mu_h\in X^*(T)$.

(b) Let (E,Φ) be a CM-type, and let $T=(\mathbb{G}_m)_{E/\mathbb{Q}}$. Then $T(\mathbb{R})=(E\otimes_{\mathbb{Q}}\mathbb{R})^{\times}\simeq(\mathbb{C}^{\Phi})^{\times}$, and we define $h_{\Phi}\colon\mathbb{S}\to T_{\mathbb{R}}$ to be the homomorphism such that $h_{\Phi}(\mathbb{R})$ is

$$z\mapsto(z,\ldots,z)\colon\mathbb{C}^{\times}\to(\mathbb{C}^{\Phi})^{\times}.$$

On $\mathbb{C}$-points, $(h_{\Phi})_{\mathbb{C}}\colon\mathbb{S}_{\mathbb{C}}\to T_{\mathbb{C}}$ is the map

$$(z_1,z_2)\mapsto(z_1,\ldots,z_1,z_2,\ldots,z_2)\colon\mathbb{C}^{\times}\times\mathbb{C}^{\times}\longrightarrow(\mathbb{C}^{\Phi})^{\times}\times(\mathbb{C}^{\bar{\Phi}})^{\times},$$

and the corresponding cocharacter μ_{Φ} is

$$z\mapsto(z,\ldots,z,1,\ldots,1)\colon\mathbb{C}^{\times}\longrightarrow(\mathbb{C}^{\Phi})^{\times}\times(\mathbb{C}^{\bar{\Phi}})^{\times}.$$

Clearly, μ_{Φ} is defined over $\mathbb{Q}^{\mathrm{a}}$, and the elements of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ fixing it are those stabilizing Φ. Therefore $E(T,\{h_{\Phi}\})$ is equal to the reflex field of (E,Φ) (see 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 1971c, 6.1). [This is a motivating example. The σ fixing the reflex field are exactly those that preserve the family of abelian varieties (with additional structure) parametrized by the Shimura variety. In other words, the reflex field is the natural field of definition of the moduli problem.]

(d) Let (G,X) be the Shimura datum defined by a quaternion algebra B over a totally real number field F, as in Example 5.24. With the notation of that example, the class $c(X)$ contains the cocharacter $\mu\colon\mathbb{G}_{m\mathbb{C}}\to G_{\mathbb{C}}$,

$$z\mapsto(1,\ldots,1)\times\left(\left(\begin{smallmatrix}z&0\\0&1\end{smallmatrix}\right),\ldots,\left(\begin{smallmatrix}z&0\\0&1\end{smallmatrix}\right)\right)\in\mathrm{GL}_{2\mathbb{C}}^{I_c}\times\mathrm{GL}_{2\mathbb{C}}^{I_{nc}}.$$

This is defined over $\mathbb{Q}^{\mathrm{a}}$, and $E(G,X)$ is the fixed field of the subgroup of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ stabilizing $I_{\mathrm{nc}}\subset I$. For example, if I_{nc} consists of a single element v, so $\mathrm{Sh}(G,X)$ is a curve, then $E(G,X)=v(F)$.

(e) Let (G,X) be a Shimura datum in which G is adjoint. Choose a maximal torus T in $G_{\mathbb{Q}^{\mathrm{a}}}$ and a base Δ for the roots of $(G,T)_{\mathbb{Q}^{\mathrm{a}}}$. Recall that the nodes of the Dynkin diagram of (G,T) are indexed by Δ. There is a natural "$*$"-action of the Galois group $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ on Δ (see A.8). Each $c \in \mathcal{C}(\mathbb{Q}^{\mathrm{a}})$ contains a unique $\mu\colon \mathbb{G}_m \to G_{\mathbb{Q}^{\mathrm{a}}}$ such that $\langle \alpha, \mu \rangle \geq 0$ for all $\alpha \in \Delta$ (because the Weyl group acts simply transitively on the Weyl chambers), and the map

$$c \mapsto (\langle \alpha, \mu \rangle)_{\alpha \in \Delta}\colon \mathcal{C}(\mathbb{Q}^{\mathrm{a}}) \to \mathbb{N}^{\Delta} \qquad \text{(product of copies of } \mathbb{N} \text{ indexed by } \Delta)$$

is bijective. Therefore, $E(G,X)$ is the fixed field of the subgroup of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ fixing $(\langle \alpha, \mu \rangle)_{\alpha \in \Delta} \in \mathbb{N}^{\Delta}$. See Deligne 1971b, p. 139.

As $G_{\mathbb{R}}$ contains a compact maximal torus, complex conjugation acts as the opposition involution on Δ. It follows that $E(G,X)$ is a totally real field if the opposition involution fixes $(\langle \alpha, \mu \rangle)_{\alpha \in \Delta}$ and otherwise it is a CM-field.

(f) Let (G,X) be a Shimura datum. Let $G_1, \ldots, G_g$ be the simple factors of G^{ad}, and let $T = G/G^{\mathrm{der}}$. Then (G,X) defines Shimura data $(G_1,X_1), \ldots, (G_r,X_r)$,and (T,h), and

$$E(G,X) = E(G_1,X_1) \cdot \cdots \cdot E(G_r,X_r) \cdot E(T,h).$$

It follows that if T is split by a CM-field (SV6), then $E(G,X)$ is either a CM-field or is totally real. See Deligne 1971b, 3.8.

Special points

DEFINITION 12.5. A point $x \in X$ is said to be ***special*** if there exists a torus[69] $T \subset G$ 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***.[70]

REMARK 12.6. If (T,x) is special, then $T(\mathbb{R})$ fixes x. Conversely, 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 $t \in T(\mathbb{R})$. Then $h(\mathbb{C}^{\times})$ is contained in the centralizer of $T(\mathbb{R})$ in $G(\mathbb{R})$. 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.

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 $(\mathbb{G}_m)_{E/\mathbb{Q}} \subset G$.

The $\mathbb{C}$-vector space $E \otimes \mathbb{C}$ has basis $\{1 \otimes 1, 1 \otimes (-z)\}$. The kernel of the map $e \otimes z \to ez\colon E \otimes \mathbb{C} \to \mathbb{C}$ is the one-dimensional space spanned by $z \otimes 1 + 1 \otimes (-z)$, which, with respect to our basis, is $\binom{z}{1}$. This represents the point $z \in \mathcal{H}_1^{\pm}$. The map is $E \otimes \mathbb{R}$-linear, and so $(E \otimes \mathbb{R})^{\times}$ fixes z, i.e., $(\mathbb{G}_m)_{E/\mathbb{Q}}(\mathbb{R})$ 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 $\mathbb{Q}$. Thus, the special points are exactly the points such that the elliptic curve $\mathbb{C}/\mathbb{Z} + \mathbb{Z}z$ has complex multiplication.

[69] Meaning, of course, defined over $\mathbb{Q}$.

[70] Because then the homomorphism $h_x\colon \mathbb{S} \to T$ factors through the Serre group, and for any representation (V,ρ) of T, $(V, \rho_{\mathbb{R}} \circ h_x)$ is the Hodge structure of a CM-motive. [Reference to be added.]

More generally, whenever the Shimura variety is a moduli variety for abelian varieties with additional structure in some natural way, the special points correspond to abelian varieties of CM-type. Therefore the theory of complex multiplication describes how an open subgroup of Aut($\mathbb{C}$) acts on the abelian variety and its additional structure, and hence on the corresponding points of the Shimura variety. In the next two subsections, we define an action of an open subgroup of Aut($\mathbb{C}$) on the special points of a Shimura variety. When the Shimura variety is a moduli variety, this agrees with the action given by complex multiplication.

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 $\sum_{\rho: E\to\mathbb{Q}^{\mathrm{a}}} \rho(Q)$ of $T(\mathbb{Q}^{\mathrm{a}})$ is stable under $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ and hence lies in $T(\mathbb{Q})$. Let $r(T,\mu)$ be the homomorphism $(\mathbb{G}_m)_{E/\mathbb{Q}} \to T$ such that

$$r(T,\mu)(P) = \sum_{\rho: E\to\mathbb{Q}^{\mathrm{a}}} \rho(\mu(P)), \quad \text{all } P \in E^{\times}. \tag{60}$$

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

$$\mathbb{A}^{\times}_{E(x)} \xrightarrow{r(T,\mu)} T(\mathbb{A}_{\mathbb{Q}}) \xrightarrow{\text{project}} T(\mathbb{A}_{\mathbb{Q},f}). \tag{61}$$

Let $a \in \mathbb{A}^{\times}_{E(x)}$, and write $a = (a_\infty, a_f) \in (E(x)\otimes_{\mathbb{Q}}\mathbb{R})^{\times} \times \mathbb{A}^{\times}_{E(x),f}$; then

$$r_x(a) = \sum_{\rho: E\to\mathbb{Q}^{\mathrm{a}}} \rho(\mu_x(a_f)).$$

Definition of a canonical model

For a special pair $(T,x) \subset (G,X)$, we have homomorphisms ((59),(61)),

$$\begin{aligned} \mathrm{art}_{E(x)} &: \mathbb{A}^{\times}_{E(x)} \twoheadrightarrow \mathrm{Gal}(E(x)^{\mathrm{ab}}/E(x)) \\ r_x &: \mathbb{A}^{\times}_{E(x)} \to T(\mathbb{A}_f). \end{aligned}$$

We write $[x,a]_K$ for the element of

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

represented by $(x,a) \in X \times G(\mathbb{A}_f)$.

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

$$\sigma[x,a]_K = [x, r_x(s)a]_K, \tag{62}$$

for all[71]

$$\left.\begin{array}{l} \sigma \in \mathrm{Gal}(E(x)^{\mathrm{ab}}/E(x)) \\ s \in \mathbb{A}^{\times}_{E(x)} \end{array}\right\} \text{ with } \mathrm{art}_{E(x)}(s) = \sigma.$$

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 (62), where s is any idèle such that $\mathrm{art}_{E(x)}(s) = \sigma|E(x)^{\mathrm{ab}}$.

[71] If $q \in G(\mathbb{Q})$ and $qx = x$, then $[x,qa]_K = [x,a]_K$, and so, according to (62), we should have $[x,r_x(s)qa]_K = [x,r_x(s)a]_K$. Following Deligne 1979, 2.2.4, I leave it to the reader to check this.

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 (62) 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 $V \rightsquigarrow V(k^{\mathrm{a}})$ is an equivalence from the category of zero-dimensional varieties over k to the category of finite sets endowed with a continuous action of $\mathrm{Gal}(k^{\mathrm{a}}/k)$. “Continuous” here just means that the action factors through $\mathrm{Gal}(L/k)$ for some finite Galois extension L of k contained in k^{a}. 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 $\mathbb{Q}^{\mathrm{a}}$, and to give a model of V over a number field E amounts to giving a continuous action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ 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 $E \overset{\text{def}}{=} E(T, h)$ 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 (62) 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 $T(\mathbb{A}_f)$. I claim that 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 prove the claim, let V be a one-dimensional E-vector space, and regard it as a $\mathbb{Q}$-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-isomorphism $a\colon H_1(A, \mathbb{Q}) \to V$ 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 $(E \otimes \mathbb{A}_f)^\times = T(\mathbb{A}_f)$. The map $(A, i, \eta) \mapsto [g]$ gives the bijection.

In (10.3) and its proof, we showed that the functor $(A, i) \rightsquigarrow (A_{\mathbb{C}}, i_{\mathbb{C}})$ defines an equivalence from the category of abelian varieties over $\mathbb{Q}^{\mathrm{a}}$ 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, $\mathrm{Sh}_K(T, h_\Phi)$ classifies isomorphism classes of triples $(A, i, \eta K)$, where (A, i) is now an abelian variety over $\mathbb{Q}^{\mathrm{a}}$ of CM-type (E, Φ).

The group $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/E^*)$ acts on the set $\mathcal{M}_K$ of such triples: let $(A, i, \eta) \in \mathcal{M}_K$; for $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/E^*)$, define $\sigma(A, i, \eta K)$ to be the triple $(\sigma A, {}^\sigma i, {}^\sigma\eta K)$, where ${}^\sigma\eta$ is the composite

$$V(\mathbb{A}_f) \xrightarrow{\eta} V_f(A) \xrightarrow{\sigma} V_f(\sigma A); \tag{63}$$

because σ fixes E^*, $(\sigma A, \sigma i)$ is again of CM-type (E, Φ).

The group $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/E^*)$ acts on $\mathrm{Sh}_K(T, h_\Phi)$ by the rule (62):

$$\sigma[g] = [r_{h_\Phi}(s)g]_K, \quad \mathrm{art}_{E^*}(s) = \sigma|E^*.$$

This defines a model of $\mathrm{Sh}_K(T, h_\Phi)$ over E^*, and next proposition shows that it is canonical.

PROPOSITION 12.11. *The map* $(A, i, \eta) \mapsto [a \circ \eta]_K : \mathcal{M}_K \to \mathrm{Sh}_K(T^E, h_\Phi)$ *commutes with the actions of* $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/E^*)$.

PROOF. Let $(A, i, \eta) \in \mathcal{M}_K$ map to $[a \circ \eta]_K$ for an appropriate isomorphism $a : H_1(A, \mathbb{Q}) \to V$, and let $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/E^*)$. According to the main theorem of complex multiplication (11.2), there exists an E-linear isogeny $\alpha : A \to \sigma A$ 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 $\sigma(A, i, \eta) \mapsto [a \circ V_f(\alpha)^{-1} \circ \sigma \circ \eta]_K$. But

$$V_f(\alpha)^{-1} \circ \sigma = N_{\Phi^*}(s) = r_{h_\Phi}(s),$$

and so

$$[a \circ V_f(\alpha)^{-1} \circ \sigma \circ \eta]_K = [r_{h_\Phi}(s) \cdot (a \circ \eta)]_K$$

as required. □

ASIDE. Every elliptic modular curve is defined in a natural way over a number field k (which depends on the curve). For analysts, the explanation for this is that the Fourier expansions at the cusps provide a k-structure on the spaces of modular forms (hence a projective embedding over k). For algebraic geometers, the explanation is that the curve is the solution of a moduli problem that is defined over k. Shimura (1967) showed that every quotient of $\mathcal{H}_1$ by a quaternionic congruence group has a "naturally-defined" model over a specific number field, even when the quotient is compact (hence without cusps) and is not a moduli variety in any natural way. This surprised both the analysts and the algebraic geometers. Although Shimura's construction of the model involved choices, he showed that it is uniquely determined by certain reciprocity laws at the special points. This was the birth of the theory of canonical models. The conjecture that every Shimura variety has a canonical model is often called Shimura's conjecture. See Appendix C.

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, p. 269, line 10, and in 2.6.3, p. 284, line 21). See my letter to Deligne, 28.03.90 (available on my website under articles).

13 Uniqueness of canonical models

In this section, we 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*

$$V \rightsquigarrow V_\Omega + \text{action of } \mathrm{Aut}(\Omega/k) \text{ on } V(\Omega)$$

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. Let $\mathfrak{g}$ be a Lie algebra over an algebraically closed field k of characteristic zero. A subalgebra $\mathfrak{h}$ is Cartan if it is nilpotent and equal to its own normalizer. When $\mathfrak{g}$ is the Lie algebra of a semisimple algebraic group G over k, then the Cartan subalgebras are exactly the Lie algebras of maximal tori in G. Let $P_x(T)$ denote the characteristic polynomial $\det(T - \mathrm{ad}(x))$ of an element x in $\mathfrak{g}$. Then we can write

$$P_x(T) = T^n + a_{n-1}(x)T^{n-1} + \cdots + a_r(x)T^r, \quad n = \dim(\mathfrak{g}),$$

with the a_i regular functions on $\mathfrak{g}$ and $a_r \neq 0$. The $x \in \mathfrak{g}$ such that $a_r(x) \neq 0$ are said to be regular. The regular x form a connected dense open subset of $\mathfrak{g}$ for the Zariski topology. The Cartan subalgebras are exactly the centralizers of regular elements of $\mathfrak{g}$, and any two are conjugate by an inner automorphism. More generally, an element of a Lie algebra over a nonalgebraically closed field k is said to be regular if it becomes so over the algebraic closure of k. See Serre 1966, Chap. III.

Now let (G, X) be a Shimura datum. 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 in $G_\mathbb{R}$ of a regular element λ of $\mathrm{Lie}(G_\mathbb{R})$. If $\lambda_0 \in \mathrm{Lie}(G)$ is chosen to be sufficiently close to λ in $\mathrm{Lie}(G_\mathbb{R})$, then it will also be regular, and so its centralizer T_0 in G is a maximal torus in G. Moreover, T_0 will become conjugate to T over $\mathbb{R}$:[72]

$$T_{0\mathbb{R}} = gTg^{-1} \text{ for some } g \in G(\mathbb{R}).$$

[72]Not all maximal tori in $G_\mathbb{R}$ are conjugate. Rather, the maximal tori fall into several connected components with any two in the same connected component being conjugate. This explains why $T_{0\mathbb{R}}$ is conjugate to T if λ_0 is chosen sufficiently close to λ.

Now $h_{gx}(\mathbb{S}) \stackrel{\text{def}}{=} ghg^{-1}(\mathbb{S}) \subset T_{0\mathbb{R}}$, and so gx is special. □

KEY LEMMA 13.4. *For every 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 1971b, 5.1. The basic idea is the same as that of the proof of 13.3 above, but the proof requires the Hilbert irreducibility theorem. □

When $G = \mathrm{GL}_2$, the key lemma just says that, for every 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. The statement of the key lemma is unsurprising, but the proof is a little delicate.

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. Then $G(\mathbb{Q})x \times G(\mathbb{A}_f)$ is dense in $X \times G(\mathbb{A}_f)$, and its image in $\mathrm{Sh}_K(G,X)(\mathbb{C})$ is dense for the complex topology and, a fortiori, for the Zariski topology. That image equals $\{[x,a]_K \mid a \in G(\mathbb{A}_f)\}$ because $[gx,b]_K = [x,g^{-1}b]_K$ for $g \in G(\mathbb{Q})$ and $b \in G(\mathbb{A}_f)$. □

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 $\mathcal{T}(g)$

$$[x,a]_K \mapsto [x,ag]_{K'}\colon \mathrm{Sh}_K(\mathbb{C}) \to \mathrm{Sh}_{K'}(\mathbb{C})$$

is well-defined. Here $\mathrm{Sh}_K = \mathrm{Sh}_K(G,X)$. The map $\mathcal{T}(g)$ is a morphism of algebraic varieties over $\mathbb{C}$ (because of Theorem 3.14).

THEOREM 13.6. *If $\mathrm{Sh}_K(G,X)$ and $\mathrm{Sh}_{K'}(G,X)$ 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 $\mathrm{art}(s) = \sigma|E(x_0)^{\mathrm{ab}}$. For $a \in G(\mathbb{A}_f)$,

$$\begin{array}{ccc}
[x_0,a]_K & \xrightarrow{\mathcal{T}(g)} & [x_0,ag]_{K'} \\
\downarrow \sigma & & \downarrow \sigma \\
[x_0,r_{x_0}(s)a]_K & \xrightarrow{\mathcal{T}(g)} & [x_0,r_{x_0}(s)ag]_{K'}
\end{array}$$

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 $\mathrm{Aut}(\mathbb{C}/E(G,X))$. □

THEOREM 13.7. *(a) A canonical model of* $\mathrm{Sh}_K(G,X)$ *(if it exists) is unique up to a unique isomorphism.*

(b) If, for all compact open subgroups K *of* $G(\mathbb{A}_f)$, $\mathrm{Sh}_K(G,X)$ *has a canonical model, then so also does* $\mathrm{Sh}(G,X)$*, and it is unique up to a unique isomorphism.*

PROOF. (a) Take $K = K'$ and $g = 1$ in Theorem 13.6. In more detail, let $(M_K(G,X),\varphi)$ and $(M'_K(G,X),\varphi')$ be canonical models of $\mathrm{Sh}_K(G,X)$ over $E(G,X)$. Then the composite

$$M_K(G,X)_{\mathbb{C}} \xrightarrow{\varphi} \mathrm{Sh}_K(G,X) \xrightarrow{\varphi'^{-1}} M'_K(G,X)_{\mathbb{C}}$$

is fixed by all automorphisms of $\mathbb{C}$ fixing $E(G,X)$, and is therefore defined over $E(G,X)$.

(b) Obvious from (13.6). □

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 $\mathrm{Sh}(G,X)$ and $\mathrm{Sh}(G',X')$ have canonical models $M(G,X)$ and $M(G',X')$. Then the morphism $\mathrm{Sh}(a)\colon \mathrm{Sh}(G,X) \to \mathrm{Sh}(G',X')$ is defined over $E(G,X)\cdot E(G',X')$.

The Galois action on the connected components

A canonical model for $\mathrm{Sh}_K(G,X)$ will define an action of $\mathrm{Aut}(\mathbb{C}/E(G,X))$ on the set $\pi_0(\mathrm{Sh}_K(G,X))$. In the case that G^{der} is simply connected, we saw in §5 that

$$\pi_0(\mathrm{Sh}_K(G,X)) \simeq T(\mathbb{Q})\backslash Y \times T(\mathbb{A}_f)/\nu(K)$$

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

$$r = r(T,\mu_h)\colon \mathbb{A}^{\times}_{E(G,X)} \to T(\mathbb{A}_{\mathbb{Q}}).$$

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[73]

$$\sigma[y,a]_K = [r(s)_\infty y, r(s)_f \cdot a]_K, \text{ for all } y \in Y, \quad a \in T(\mathbb{A}_f). \tag{64}$$

When we use (64) 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 (64) follows from (62), 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 1971b, §3, except that I am more unscrupulous in my use of Zorn's lemma.

[73]To prove that this description is correct, check that, for each special $x_0 \in X$, the map $\mathrm{Sh}_K(\mathbb{C}) \to T(\mathbb{Q})\backslash Y \times T(\mathbb{A}_f)/\nu(K)$ is $\mathrm{Aut}(\mathbb{C}/E(x_0))$-equivariant on the points $[x_0,a]_K$ for $a \in G(\mathbb{A}_f)$, and then apply 13.4.

14 Existence of canonical models

Canonical models are known to exist for all Shimura varieties. In this section, we 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

$$\{\text{varieties over } k\} \rightsquigarrow \{\text{varieties } V \text{ over } \Omega + \text{action of } \mathcal{A} \text{ on } V(\Omega)\},$$

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 σP is the element of $(\sigma V)(\Omega)$ 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

$$\sigma P \mapsto \sigma \cdot P : (\sigma V)(\Omega) \to V(\Omega)$$

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 $f_\sigma : \sigma V \to V$. 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 1956a (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 exist points $P_1,\ldots,P_n$ in $V(\Omega)$ satisfying the conditions (a) and (b) of (14.4).*

PROOF. Combine 14.4 and 14.5. □

For an elementary proof of the corollary, not using the results of Weil 1956a, see AG 16.33.

Review of local systems and families of abelian varieties

Let S be a connected 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 $\gamma\colon [0,1] \to S$ be a path in S. The interval $[0,1]$ is simply connected, and there is a unique isomorphism from γ^*F to the constant sheaf defined by $F_{\gamma(0)}$ that restricts to the identity map on the fibres at 0. This isomorphism defines an isomorphism between $(\gamma^*F)_0$ and $(\gamma^*F)_1$, i.e., an isomorphism

$$\gamma(F)\colon F_{\gamma(0)} \to F_{\gamma(1)}.$$

This isomorphism depends only on the homotopy class of γ, and satisfies the condition $\gamma\gamma'(F) = \gamma(F)\cdot\gamma'(F)$. Therefore, for every $o \in S$, we obtain an action ρ_o of the fundamental group $\pi_1(S,o)$ on the fibre F_o at o. The proof of the next statement is well known and easy (but not easy to find in the literature).

PROPOSITION 14.7. *If S is connected, then $F \rightsquigarrow (F_o, \rho_o)$ is an equivalence from the category of local systems of $\mathbb{Z}$-modules on S to the category of free $\mathbb{Z}$-modules of finite rank endowed with an action of $\pi_1(S, o)$.*

In particular, if S is simply connected, then every local system on S is trivial.

Now let S be a complex manifold, and let F be a local system of $\mathbb{Z}$-modules on S. 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 $(F \otimes \mathbb{R}, (h_s))$ becomes a variation of Hodge structures in the sense of §2 on every open subset of S on which the local system F is trivial. This is equivalent to requiring that the pull-back of $(F \otimes \mathbb{R}, (h_s))$ to the universal covering of S is a local system in the sense of §2. A ***polarization*** of a variation of Hodge structures $(F, (h_s))$ is a pairing $\psi: F \times F \to \mathbb{Z}$ 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 $f: A \to V$ 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 $\mathbb{C}$. Then $(A, f) \rightsquigarrow (R^1 f_* \mathbb{Z})^\vee$ is an equivalence 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 1971a, 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 (see §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 $V(\mathbb{C})$:

$$V(\mathbb{C}) = L \oplus L', \quad L, L' \text{ totally isotropic.} \tag{65}$$

Every symplectic basis for $V(\mathbb{C})$ defines such a pair, and the 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 p. 111). 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*** (or of ***CM-type***) 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 write E is a product of CM-fields, $E = E_1 \times \cdots \times E_m$. 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).[74]

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. The statements depend only on A up to isogeny, and every abelian variety is isogenous to a product of simple abelian varieties. It follows that we may assume that A is simple. For a simple A, we shall prove that the following conditions are equivalent:

(a) A is of CM-type;

(b) $\mathrm{End}^\circ(A)$ is a CM-field of degree $2\dim A$ over $\mathbb{Q}$;

(c) there exists a torus $T \subset \mathrm{GL}(H_1(A, \mathbb{Q}))$ such that $h_A(\mathbb{C}^\times) \subset T(\mathbb{R})$;

(d) there exists a torus $T \subset \mathrm{GL}(H_1(A, \mathbb{Q}))$ such that $\mu_A(\mathbb{C}^\times) \subset T(\mathbb{C})$.

The equivalences (a)$\Leftrightarrow$(b) and (c)$\Leftrightarrow$(d) are easy (the second doesn't use that A is simple). As A is simple, $E = \mathrm{End}^0(A)$ is a division algebra of degree $\leq \dim H_1(A, \mathbb{Q})$ over $\mathbb{Q}$.

(b)$\Rightarrow$(c). Let A be an abelian variety such that $\mathrm{End}^0(A)$ contains a field E for which $H_1(A, \mathbb{Q})$ has dimension 1 as an E-vector space. The action of $E \otimes \mathbb{R}$ on $H_1(A, \mathbb{R})$ preserves the Hodge structure, and so $h_A(\mathbb{C}^\times)$ commutes with $E \otimes \mathbb{R}$ in $\mathrm{End}(H_1(A, \mathbb{R}))$. Therefore

$$h_A(\mathbb{C}^\times) \subset (E \otimes \mathbb{R})^\times = (\mathbb{G}_m)_{E/\mathbb{Q}}(\mathbb{R}).$$

(d)$\Rightarrow$(b). As for any abelian variety, $\mathrm{End}^0(A)$ is the subalgebra of $\mathrm{End}(H_1(A, \mathbb{Q}))$ of elements preserving the Hodge structure or, equivalently, that commute with $\mu_A(\mathbb{G}_m)$ in $\mathrm{GL}(H_1(A, \mathbb{C}))$. By assumption, there is a torus $T \subset \mathrm{GL}(H_1(A, \mathbb{Q}))$ such that $\mu_A(\mathbb{C}^\times) \subset T(\mathbb{C})$. Therefore

$$\mathrm{End}^0(A) \supset \{\alpha \in \mathrm{End}(H_1) \mid \alpha \text{ commutes with the action of } T\}$$

and so

$$\begin{aligned}\mathrm{End}^0(A) \otimes \mathbb{C} &\supset \{\alpha \in \mathrm{End}(H_1) \mid \alpha \text{ commutes with the action of } T\} \otimes \mathbb{C} \\ &= \{\alpha \in \mathrm{End}(H_1 \otimes \mathbb{C}) \mid \alpha \text{ commutes with the action of } T_{\mathbb{C}}\}.\end{aligned}$$

[74] If $A(\mathbb{C}) = \mathbb{C}^g / \Lambda$, then

$$H_1(A, \mathbb{Z}) = \Lambda, \quad H_1(A, \mathbb{Q}) = \Lambda \otimes \mathbb{Q}, \quad H_1(A, \mathbb{R}) = \Lambda \otimes \mathbb{R} \simeq \mathbb{C}^g.$$

Because T is a torus, $H_1(A,\mathbb{C}) = \bigoplus_{\chi\in X^*(T)} H_\chi$, and so $\mathrm{End}_T(H_1\otimes\mathbb{C})$ contains an étale $\mathbb{C}$-algebra of degree $2\dim A$. It follows that $\mathrm{End}^0(A)$ does also.

It remains to show that $E \overset{\text{def}}{=} \mathrm{End}^0(A)$ is a CM-field. We shall show that the involution $*$ of E defined by any Riemann form on $H_1(A,\mathbb{Q})$ is a complex conjugation on E, i.e., a nontrivial involution such that $\rho(a^*) = \overline{\rho(a)}$ for every embedding $\rho\colon E\hookrightarrow\mathbb{C}$. Note that $h(i)\in E\otimes\mathbb{R}$. Let ψ be a Riemann form corresponding to some polarization on A. The Rosati involution $e\mapsto e^*$ on E is determined by the condition

$$\psi(x,ey) = \psi(e^*x,y), \quad x,y\in H_1(A,\mathbb{Q}).$$

It follows from

$$\psi(x,y) = \psi(h(i)x,h(i)y)$$

that

$$h(i)^* = h(i)^{-1} \quad (= -h(i)).$$

The Rosati involution therefore is nontrivial on E, and E has degree 2 over its fixed field F. There exists an $\alpha\in F^\times$ such that

$$E = F[\sqrt{\alpha}], \quad \sqrt{\alpha}^* = -\sqrt{\alpha},$$

and α is uniquely determined up to multiplication by a square in F. If E is identified with $H_1(A,\mathbb{R})$ through the choice of an appropriate basis vector, then

$$\psi(x,y) = \mathrm{Tr}_{E/\mathbb{Q}}\,\alpha x y^*, \quad x,y\in E,$$

(cf. A.7). The positivity condition on ψ implies that

$$\mathrm{Tr}_{E\otimes\mathbb{R}/\mathbb{R}}(fx^2) > 0, \quad x\neq 0, \quad x\in F\otimes\mathbb{R}, \quad f = \alpha/h(i),$$

which implies that F is totally real. Moreover, for every embedding $\sigma\colon F\hookrightarrow\mathbb{R}$, we must have $\sigma(\alpha)<0$, for otherwise $E\otimes_{F,\sigma}\mathbb{R} = \mathbb{R}\times\mathbb{R}$ with $(r_1,r_2)^* = (r_2,r_1)$, and the positivity condition is impossible. Thus, $\sigma(\alpha)<0$, and $*$ is complex conjugation relative to any embedding of E into $\mathbb{C}$. This completes the proof. □

COROLLARY 14.11. *If $(A,s,\eta K)$ maps to $[x,a]_K$ under $\mathcal{M}_K\to \mathrm{Sh}_K$, then A is CM 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, pp. 29–30) which is preserved by σ. Therefore, $\pm{}^\sigma s$ is a polarization for $H_1(\sigma A,\mathbb{Q})$. We define $\sigma(A,s,\eta K)$ to be $(\sigma A,{}^\sigma s,{}^\sigma\eta K)$ with ${}^\sigma\eta$ as in (63).

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 (62) 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. □

ASIDE. Explain the existence using Mumford's results.

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 $\mathrm{Sh}_K(\mathbb{C})$, which we denote by $\cdot$. If this action satisfies the conditions of hypotheses of Corollary 14.6, then Sh_K has a model over $\mathbb{Q}$, which Proposition 14.12 will show to be canonical.

Condition (a) of (14.6). We know that Sh_K 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 $\pi_0(\mathrm{Sh}_K) \simeq \mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$. 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 $\nu(a\circ\eta) \in \varepsilon$.

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:

$$\begin{array}{ccc} X^+ & \xleftarrow{\ \sigma\ } & U \\ \downarrow & & \downarrow \\ X^+(\Gamma_{\sigma\varepsilon}) & \xleftarrow{\ f_\sigma\ } & \sigma(X^+(\Gamma_\varepsilon)) \\ \mathcal{M}_K^{\sigma\varepsilon} & \xleftarrow{\ \sigma\ } & \mathcal{M}_K^\varepsilon \end{array}$$

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 $f:\mathcal{A}\to X^+(\Gamma_\varepsilon)$. As f is a regular map of algebraic varieties, we can apply σ to it, and obtain a polarized family of abelian varieties $\sigma f:\sigma\mathcal{A}\to\sigma(X^+(\Gamma_\varepsilon))$. 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 be 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

$$H^n(A,\mathbb{A}_f)\simeq \mathrm{Hom}(\textstyle\bigwedge^n\Lambda,\mathbb{A}_f)\simeq\mathrm{Hom}(\bigwedge^n(\Lambda\otimes\mathbb{A}_f),\mathbb{A}_f)\simeq\mathrm{Hom}(\bigwedge^n V_fA,\mathbb{A}_f)$$

(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 in 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. For a concise exposition of the proof, with some simplifications, see Milne 2013, 9.9. [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 $\mathbb{Q}^{\mathrm{a}}$ 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) $(1 \le i \le m)$ 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. A connected Shimura datum (H,X^+) is of primitive abelian type if and only if it is of the form (G^{der},X^+), where (G,X) a Shimura datum of Hodge type (this is almost the definition), and so $\mathrm{Sh}^{\circ}(H,X^+)$ has a canonical model because $\mathrm{Sh}(G,X)$ does (14.15). Now (14.16) proves the existence of canonical models 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

As noted above, Deligne proved the existence of canonical models for Shimura varieties of abelian type in his Corvallis article (Deligne 1979). At the time, Shimura's conjecture remained open for Shimura varieties of type E_6, E_7, and many of type D.[75] There was some progress using analytic methods (Baily, Karel, Garrett), but the general case seemed to be beyond reach until Piatetski-Shapiro suggested using Shimura subvarieties of type A_1. This was perhaps suggested by the fact that the standard method for studying split reductive groups is to exploit their subgroups of type A_1.

By the techniques developed by Deligne, it suffices to prove Shimura's conjecture for a connected Shimura datum (G, X^+) with G simple and simply connected. Then $G = (H)_{F/\mathbb{Q}}$, where F is a totally real field and H is geometrically simple over F. In general, $\mathrm{Sh}^\circ(G, X^+)$ will contain no Shimura subvarieties of type A_1, but it will after we have replaced G with $G_* = (H_{F'})_{F'/\mathbb{Q}}$, where F' is a totally real field containing F and large enough that $H_{F'}$ splits over CM-extension of degree 2. The idea is then to prove Shimura's conjecture for $\mathrm{Sh}^\circ(G_*, X_*^+)$ by exploiting its Shimura subvarieties of type A_1, and then deduce it for $\mathrm{Sh}^\circ(G, X^+)$ by using that it is a Shimura subvariety of $\mathrm{Sh}^\circ(G_*, X_*^+)$.

After Borovoi had unsuccessfully attempted to use this idea to prove Shimura's conjecture, the author used it to prove a conjecture of Langlands on the conjugation of Shimura varieties,[76] which has Shimura's conjecture as a consequence (Milne 1983). This completed the proof of Shimura's conjecture for all Shimura varieties.

Let (G, X) be a Shimura datum, and let σ be an automorphism of $\mathbb{C}$. Langlands defines a new Shimura datum (G^σ, X^σ) and conjectured that there exists an isomorphism

$$f_\sigma : \sigma\mathrm{Sh}(G, X) \to \mathrm{Sh}(G^\sigma, X^\sigma)$$

satisfying certain conditions sufficient to determine it uniquely.[77] The maps f_σ for σ fixing the reflex field $E(G, X)$ form a descent datum, and the descended variety is canonical. Although the conjecture is stronger than that of Shimura, it is often easier to work with because it involves only varieties over $\mathbb{C}$ (not varieties over possibly different reflex fields). As far as I know, Shimura's conjecture has not been proved except through Langlands's conjecture. This approach to proving Shimura's conjecture is independent of the previous approach, and hence of the moduli of abelian varieties, except for results on Shimura varieties defined by groups of type A_1 over totally real fields (see p. 94).

ASIDE. Langlands's conjugacy conjecture also grew out of his work on the zeta functions of Shimura varieties. Let X be an algebraic variety over a number field E. The zeta function of X has a local factor for each embedding $\sigma : E \hookrightarrow \mathbb{C}$, which depends on the geometry of σX. The reflex field E of a Shimura variety is (by definition) a subfield of $\mathbb{C}$. For the canonical model X over E, we know what σX is when σ is the given embedding, but what if it is a different embedding? The axiom of choice allows us to extend σ to an automorphism of $\mathbb{C}$, and the answer is given by Langlands's conjecture.

[75] As the dimension tends to infinity, the proportion of Shimura varieties that are of abelian type tends to zero.

[76] Langlands's conjecture had been proved earlier for Shimura varieties of abelian type in Milne and Shih 1979.

[77] More precisely, (G^σ, X^σ) and f_σ depend on the choice of an $h \in X$, but only up to a given isomorphism.

Langlands has suggested that the fact that his conjecture, arising from the study of zeta functions, permitted us to complete the proof of Shimura's conjecture is another indication that we algebraists need to allow ourselves to be guided by the analysts.

Final remark: rigidity

One might expect that if one modified the condition (62), 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 (62) is the *only* one for which canonical models can exist. 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 (62).

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, we explain how to define such a category.

The goal of this section and the next is to give an elementary statement of the conjecture of Langlands and Rapoport, and in the following section we explain how the conjecture leads to a formula that permits the analysts to compute the local zeta function. In an aside at the end of the section (p. 140), we explain the philosophy underlying the construction.

Semisimple categories

An object of an abelian category M is ***simple*** if it is nonzero and has no proper nonzero subobjects. Let F be a field. By an ***F-category***, we 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, $\mathrm{End}(re) \simeq M_r(\mathrm{End}(e))$. 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 $x = \sum r_i e_i$ $(r_i \geq 0)$ and $y = \sum s_i e_i$ $(s_i \geq 0)$ 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 our conventions: 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.

According to 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[78] 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:

$$\mathbb{H} = \mathbb{C} \oplus \mathbb{C}j, \quad j^2 = -1, \quad jzj^{-1} = \bar{z} \quad (z \in \mathbb{C}).$$

(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 $x \mapsto x^q$ on the residue field of L, where q is the size of the residue field of F. For each i, $1 \leq i \leq n$, we define an F-algebra as follows:

$$D_{i,n} = Le_0 \oplus Le_1 \oplus \cdots \oplus Le_{n-1}$$

as an F-vector space and the multiplication is determined by

$$e_j \cdot c = \sigma^j c \cdot e_j \text{ all } c \in L$$
$$e_j e_l = \begin{cases} e_{j+l} & \text{if } \quad j+l \leq n-1 \\ \pi^i e_{j+l-n} & \text{if } \quad j+l > n-1. \end{cases}$$

Identify L with a subfield of $D_{i,n}$ by identifying e_0 with 1, and let $a = e_1$; now

$$D_{i,n} = L \oplus La \oplus \cdots \oplus La^{n-1}, \quad a^n = \pi^i, \quad aca^{-1} = \sigma(c) \quad (z \in L).$$

Then $D_{i,n}$ is a central simple algebra over F, and it 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 associative, 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)

$$D \otimes_F D^{\mathrm{opp}} \simeq \mathrm{End}_{F\text{-linear}}(D) \approx M_r(F),$$

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.

[78] If F is algebraically closed, then each element of a central division algebra over F generates a field of finite degree over F, and so lies in F. For the proof in the finite case, see CFT, IV 4.1.

(b) The Brauer group $\mathbb{R}$ has order two: $\mathrm{Br}(\mathbb{R}) \simeq \frac{1}{2}\mathbb{Z}/\mathbb{Z}$.

(c) For a p-adic field F, the map $[D_{n,i}] \mapsto \frac{i}{n} \mod \mathbb{Z}$ is an isomorphism $\mathrm{Br}(F) \simeq \mathbb{Q}/\mathbb{Z}$.

(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

$$0 \to \mathrm{Br}(F) \xrightarrow{[B]\mapsto[B\otimes_F F_v]} \bigoplus_v \mathrm{Br}(F_v) \xrightarrow{\sum_v \mathrm{inv}_v} \mathbb{Q}/\mathbb{Z} \to 0.$$

is exact.

Statement (d) is shown in the course of proving the main theorem of class field theory by the cohomological approach (CFT, VIII 4.2). 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

$$\mathrm{Br}(F) \simeq H^2(F, \mathbb{G}_m) \stackrel{\text{def}}{=} H^2(\mathrm{Gal}(F^{\mathrm{s}}/F), F^{\mathrm{a}\times}) \stackrel{\text{def}}{=} \varinjlim H^2(\mathrm{Gal}(L/F), L^\times).$$

The last limit is over the fields $L \subset F^{\mathrm{s}}$ of finite degree and Galois over G. This isomorphism can be most elegantly defined as follows. Let D be a central division algebra of degree n^2 over F, and assume[79] 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 theorem of Skolem and Noether (footnote 52, p. 80) 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 E^λ 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 $(\mathbb{Q}_p[T]/(T^r - p^s)) \otimes_{\mathbb{Q}_p} L$, and if $\lambda < 0$, E^λ can be taken to be the dual of $E^{-\lambda}$. See Demazure 1972, Chap. IV.

Abelian varieties

Recall (p. 98) 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 p. 98) as its simple objects. Amazingly, when k is finite, we know its numerical invariants.

[79]This will always be true when F is a p-adic or number field, but is not true (or, at least, not known to be true) for other fields. In the general case, it becomes true after D has been replaced by $M_r(D)$ for some r.

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'.

Recall also (10.7) that the Frobenius endomorphism π_A of an abelian variety A over $\mathbb{F}_q$ lies in the centre of $\mathrm{End}^0(A)$ and is a Weil q-integer.

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 $D \overset{\text{def}}{=} \mathrm{End}^0(A)$ is $F = \mathbb{Q}[\pi_A]$, and for a prime v of F,*

$$\mathrm{inv}_v(D) = \begin{cases} \frac{1}{2} & \text{if } v \text{ is real} \\ \frac{\mathrm{ord}_v(\pi_A)}{\mathrm{ord}_v(q)}[F_v\colon\mathbb{Q}_p] & \text{if } v|p \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $2\dim A = [D\colon F]^{\frac{1}{2}} \cdot [F\colon\mathbb{Q}]$.

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 $\mathbb{Q}^{\mathrm{a}} \subset \mathbb{C}$. Then the theorem gives a bijection

$$\Sigma(\mathsf{AV}^0(\mathbb{F}_q)) \to \Gamma\backslash W_1(q), \quad \Gamma = \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q}).$$

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 1968 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 $\pi \mapsto \pi^m\colon W_1(p^n) \to W_1(p^{nm})$. 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 mth-power (obviously): $\pi_{A_{\mathbb{F}_{q^m}}} = \pi_A^m$.

Let A be a simple abelian variety defined over $\mathbb{F}$, and let A_0 be a model of A over $\mathbb{F}_q$. The above remarks show that $s_A(v) \overset{\text{def}}{=} \frac{\mathrm{ord}_v(\pi_{A_0})}{\mathrm{ord}_v(q)}$ is independent of the choice of A_0. Moreover, for any $\rho\colon\mathbb{Q}[\pi_{A_0}] \hookrightarrow \mathbb{Q}^{\mathrm{a}}$, 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 $D \overset{\text{def}}{=} \mathrm{End}^0(A)$ is isomorphic to $F = \mathbb{Q}\{\pi_A\}$, and for any prime v of F,*

$$\mathrm{inv}_v(D) = \begin{cases} \frac{1}{2} & \text{if } v \text{ is real} \\ s_A(v) \cdot [F_v\colon\mathbb{Q}_p] & \text{if } v|p \\ 0 & \text{otherwise.} \end{cases}$$

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 p. 17, to give a representation ρ of T on an F-vector space V amounts to giving an $X^*(T)$-gradation $V(L) = \bigoplus_{\chi \in X^*(T)} V_\chi$ 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 $\Gamma = \mathrm{Gal}(F^{\mathrm{a}}/F)$. 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 (Milne 2017, 12.30). □

REMARK 15.7. Let $\chi \in X^*(T)$, and let $\Gamma(\chi)$ and $F(\chi)$ be as in the proposition. Then $\mathrm{Hom}(F(\chi), F^{\mathrm{a}}) \simeq \Gamma/\Gamma(\chi)$, and so $X^*((\mathbb{G}_m)_{F(\chi)/F}) = \mathbb{Z}^{\Gamma/\Gamma(\chi)}$. The map

$$\sum\nolimits_\sigma n_\sigma \sigma \mapsto \sum\nolimits_\sigma n_\sigma \sigma\chi \colon \mathbb{Z}^{\Gamma/\Gamma(\chi)} \to X^*(T)$$

defines a homomorphism

$$T \to (\mathbb{G}_m)_{F(\chi)/F}. \tag{66}$$

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 $H^2(F,(\mathbb{G}_m)_{F(\chi)/F}) \simeq H^2(F(\chi),\mathbb{G}_m)$, which is the Brauer group of $F(\chi)$. On composing these maps, we get a homomorphism

$$H^2(F,T) \to \mathrm{Br}(F(\chi)). \tag{67}$$

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

In the remainder of this section, F is of characteristic zero. 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 up to conjugation, i.e., if e_σ in E maps to $\sigma \in \Gamma$, then there is an $a \in G(L)$ such that

$$e_\sigma g e_\sigma^{-1} = a(\sigma g)a^{-1} \quad \text{for all } g \in T(F^a).$$

For example, there is always the ***split extension*** $E_G \overset{\text{def}}{=} G(L) \rtimes \Gamma$.

An extension E is ***affine*** if its pull-back to some open subgroup of Γ is split.[80] 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 call G the ***kernel*** of the affine extension.

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

$$e_{\sigma\tau} = d_{\sigma,\tau} e_\sigma e_\tau, \quad d_{\sigma,\tau} \in T(F^{\mathrm{a}}).$$

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

$$\begin{tikzcd}
1 \arrow[r] & G_1(L) \arrow[r] \arrow[d] & E_1 \arrow[r] \arrow[d, "\phi"] & \Gamma \arrow[r] \arrow[d, equal] & 1 \\
1 \arrow[r] & G_2(L) \arrow[r] & E_2 \arrow[r] & \Gamma \arrow[r] & 1
\end{tikzcd}$$

such that the restriction of the homomorphism ϕ to $G_1(L)$ is defined by a homomorphism of algebraic groups (over L).[81] 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 $\mathsf{Rep}(G) \rightsquigarrow \mathsf{Rep}(E_G)$ is an equivalence of categories. The proof of this uses that $H^1(L/F, \mathrm{GL}(V)) = 1$.

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, $\mathrm{End}(V_{\Gamma_\chi})$ has centre $F(\chi)$, and its class in $\mathrm{Br}(F(\chi))$ is the image of $cl(E)$ under the homomorphism (67).*

---

[80] An affine extension is a down-to-earth realization of an affine groupoid. Langlands and Rapoport (1987, §1) use the term Galoisgerbe. See the aside and note at the end of this section (p. 140).

[81] As F has characteristic zero, there is at most one homomorphism of algebraic groups acting as $\phi_1|G_1(L)$ on the L-points when G_1 is connected or L is algebraically closed; otherwise we specify it as part of the data.

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 torsion-free $\mathbb{Z}$-modules with a continuous action of Γ.[82] Here continuous means that every element of the module is fixed by an open subgroup of Γ. Let $L=F^{\mathrm{a}}$. By an ***affine extension with kernel*** T, we mean an exact sequence

$$1\to T(F^{\mathrm{a}})\to E\to\Gamma\to 1$$

whose push-out

$$1\to T_i(F^{\mathrm{a}})\to E_i\to\Gamma\to 1$$

by $T(F^{\mathrm{a}})\to T_i(F^{\mathrm{a}})$ is an affine extension in the previous sense. A representation of such an extension is defined exactly as before.

REMARK 15.10. Let

$$\begin{tikzcd} L \arrow[r, dash, "\subset"] \arrow[d, dash, "\Gamma"] & L' \arrow[d, dash, "\Gamma'"] \\ F \arrow[r, dash, "\subset"] & F' \end{tikzcd}$$

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 $L_n^\times\to\mathbb{Q}_p^{\mathrm{un}\times}$,

$$\begin{tikzcd} & & & \Gamma \arrow[d] & \\ 1 \arrow[r] & L_n^\times \arrow[r] \arrow[d] & N(L_n^\times) \arrow[r] & \Gamma_n \arrow[r] & 1 \\ & \mathbb{Q}_p^{\mathrm{un}\times} & & & \end{tikzcd}$$

[82] Over a general ring R, a direct limit of finitely generated free R-modules need not be free, but it is flat. Moreover, every flat R-module arises in this way (theorem of D. Lazard). For $\mathbb{Z}$-modules, "flat" is equivalent to "torsion-free".

we obtain a $\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p$-affine extension

$$1 \to \mathbb{Q}_p^{\mathrm{un}\times} \to D_n \to \mathrm{Gal}(\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p) \to 1$$

with kernel $\mathbb{G}_m$. From a representation

$$\rho\colon D_n \to E_V = \mathrm{GL}(V(\mathbb{Q}_p^{\mathrm{un}})) \rtimes \mathrm{Gal}(\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p)$$

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 $(1,a)$ under ρ 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, $\sigma^i \mapsto a^i$ $(0 \le i \le n-1)$ 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 $\mathbb{G} \overset{\mathrm{def}}{=} \varprojlim \mathbb{G}_m$. 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 $\mathbb{Q}^{\mathrm{a}\times}$ generated by $W_1(p^n)$, and let $W = \varinjlim_n W(p^n)$.

Then W is a free $\mathbb{Z}$-module of infinite rank with a continuous action of $\Gamma = \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$. 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 for a prime v of $\mathbb{Q}\{\pi\}$ above p, we write

$$s_\pi(v) = \frac{\mathrm{ord}_v(\pi_n)}{\mathrm{ord}_v(p^n)}.$$

THEOREM 15.12. *Let P be the protorus over $\mathbb{Q}$ with $X^*(P) = W$. Then there exists an affine extension*

$$1 \to P(\mathbb{Q}^{\mathrm{a}}) \to \mathfrak{P} \to \Gamma \to 1$$

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*

$$\mathrm{inv}_v(D) = \begin{cases} \frac{wt(\pi)}{2} & \text{if } v \text{ is real} \\ s_\pi(v)\cdot[\mathbb{Q}\{\pi\}_v : \mathbb{Q}_p] & \text{if } v|p \\ 0 & \text{otherwise.} \end{cases}$$

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:

$$c \mapsto (c(\pi))\colon H^2(\mathbb{Q},P) \overset{(67)}{\to} \prod_{\Gamma\pi \in \Gamma\backslash W} \mathrm{Br}(\mathbb{Q}\{\pi\}).$$

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[83] of $\mathbb{Q}$, and choose a prime w_l of $\mathbb{Q}^{\mathrm{a}}$ dividing l. Let $\mathbb{Q}_l^{\mathrm{a}}$ be the algebraic closure of $\mathbb{Q}_l$ in the completion of $\mathbb{Q}^{\mathrm{a}}$ at w_l. Then $\Gamma_l \overset{\text{def}}{=} \mathrm{Gal}(\mathbb{Q}_l^{\mathrm{a}}/\mathbb{Q}_l)$ is a closed subgroup of $\Gamma \overset{\text{def}}{=} \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$, and we have a diagram

$$\begin{tikzcd} \mathbb{Q}^{\mathrm{a}} \arrow[r, dash] \arrow[d, dash, "\Gamma"] & \mathbb{Q}_l^{\mathrm{a}} \arrow[d, dash, "\Gamma_l"] \\ \mathbb{Q} \arrow[r, dash] & \mathbb{Q}_l. \end{tikzcd} \tag{68}$$

From $\mathfrak{P}$ we obtain a $\mathbb{Q}_l^{\mathrm{a}}/\mathbb{Q}_l$-affine extension $\mathfrak{P}(l)$ by pulling back by $\Gamma_l \to \Gamma$ and pushing out by $P(\mathbb{Q}^{\mathrm{a}}) \to P(\mathbb{Q}_l^{\mathrm{a}})$ (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*

$$\begin{tikzcd} & & & \Gamma_\ell \arrow[dl, "\zeta_\ell"'] \arrow[d, equal] & \\ 1 \arrow[r] & P(\mathbb{Q}_\ell^{\mathrm{a}}) \arrow[r] & \mathfrak{P}(\ell) \arrow[r] & \Gamma_\ell \arrow[r] & 1 \end{tikzcd}$$

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. On pulling back by $\Gamma_\ell \to \Gamma$ and pushing out using the commutative diagram

$$\begin{tikzcd} P(\mathbb{Q}^{\mathrm{a}}) \arrow[r] \arrow[d] & P(\mathbb{Q}_\ell^{\mathrm{a}}) \arrow[d] \\ \mathrm{GL}(V(\mathbb{Q}^{\mathrm{a}})) \arrow[r] & \mathrm{GL}(V(\mathbb{Q}_\ell^{\mathrm{a}})), \end{tikzcd}$$

we obtain from ρ a representation

$$\rho(\ell)\colon \mathfrak{P}(\ell) \to \mathrm{GL}(V(\mathbb{Q}_\ell^{\mathrm{a}})) \rtimes \Gamma_\ell$$

of $\mathfrak{P}(\ell)$ on $V(\mathbb{Q}_\ell)$.

[83] So $l \in \{2, 3, 5, \ldots, \infty\}$. The symbol ℓ always denotes a prime $\neq p, \infty$.

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 $V(\mathbb{Q}_\ell^{\mathrm{a}})$ 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 $V(\mathbb{Q}_\ell^{\mathrm{a}})$, which one can check to be continuous. Therefore (AG, 16.7), $V_\ell(\rho) \stackrel{\text{def}}{=} V(\mathbb{Q}_\ell^{\mathrm{a}})^{\Gamma_\ell}$ is a $\mathbb{Q}_\ell$-structure on $V(\mathbb{Q}_\ell^{\mathrm{a}})$. In this way, we get a functor $\rho \rightsquigarrow V_\ell(\rho)$ from the category of fake motives over $\mathbb{F}$ to vector spaces over $\mathbb{Q}_\ell$.

With a little more effort, it is possible to define a functor

$$\rho \rightsquigarrow V_f^p(\rho) \quad (\text{free module over } \mathbb{A}_f^p \stackrel{\text{def}}{=} \prod_{\ell \neq p, \infty} (\mathbb{Q}_\ell, \mathbb{Z}_\ell))$$

such that $V_\ell(\rho) = V_f^p(\rho) \otimes_{\mathbb{A}_f^p} \mathbb{Q}_\ell$ for all $\ell \neq p, \infty$.

THE ISOCRYSTAL OF A FAKE MOTIVE

Choose a prime w_p of $\mathbb{Q}^{\mathrm{a}}$ dividing p, and let $\mathbb{Q}_p^{\mathrm{un}}$ and $\mathbb{Q}_p^{\mathrm{a}}$ denote the subfields of the completion of $\mathbb{Q}^{\mathrm{a}}$ at w_p. Then $\Gamma_p \stackrel{\text{def}}{=} \mathrm{Gal}(\mathbb{Q}_p^{\mathrm{a}}/\mathbb{Q}_p)$ is a closed subgroup of $\Gamma \stackrel{\text{def}}{=} \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ and $\Gamma_p^{\mathrm{un}} \stackrel{\text{def}}{=} \mathrm{Gal}(\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p)$ 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 $H^2(\Gamma_p, P(\mathbb{Q}_p^{\mathrm{a}}))$ arises from a cohomology class in $H^2(\Gamma_p^{\mathrm{un}}, P(\mathbb{Q}_p^{\mathrm{un}}))$ (proof omitted).

(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 $\mathfrak{P}(p)^{\mathrm{un}}$. □

In summary, we have a diagram

$$\begin{tikzcd}
1 \arrow[r] & \mathbb{G}_m(\mathbb{Q}_p^{\mathrm{un}}) \arrow[r] \arrow[d] & D \arrow[r] \arrow[d] & \Gamma_p^{\mathrm{un}} \arrow[r] \arrow[d, equal] & 1 \\
1 \arrow[r] & P(\mathbb{Q}_p^{\mathrm{un}}) \arrow[r] \arrow[d] & \mathfrak{P}(p)^{\mathrm{un}} \arrow[r] \arrow[d, dash, dashed] & \Gamma_p^{\mathrm{un}} \arrow[r] & 1 \\
1 \arrow[r] & P(\mathbb{Q}_p^{\mathrm{a}}) \arrow[r] & \mathfrak{P}(p) \arrow[r] & \Gamma_p \arrow[r] \arrow[u] & 1.
\end{tikzcd}$$

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 following 15.13). 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 $\mathbb{Q}^{\mathrm{a}}$ 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,*

$$\phi_\mu : \mathfrak{P} \to E_T.$$

PROOF. This is obvious from the fact, implicit in the statement following 15.12, that the functor sending an object of $\mathsf{AV}^0(\mathbb{F})$ to the corresponding fake abelian variety is an equivalence of categories. □

Let A be an abelian variety of CM-type (E, Φ) over $\mathbb{Q}^{\mathrm{a}}$, 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 $\phi : \mathfrak{P} \to E_T$. 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 $\mathsf{AV}(\mathbb{F})$.[84]

ASIDE. Over every field k, there is a $\mathbb{Q}$-linear semisimple category $\mathsf{Mot}(k)$ of numerical motives. Now assume k to be the algebraic closure of $\mathbb{F}_p$. When one assumes that the Tate conjecture and Grothendieck's standard conjectures hold over k, then it is possible to describe $\mathsf{Mot}(k)$ quite explicitly: in particular, the isomorphism classes of objects are indexed by the elements of $\Gamma \backslash W$ and the endomorphism algebras have the same description as in Theorem 15.12(b) (folklore; see Milne 1994a). The category $\mathsf{Mot}(k)$ is Tannakian. If it had a fibre functor over $\mathbb{Q}$, then the choice of such a functor would identify the category with the category of representations of an affine group scheme. The category $\mathsf{Mot}(k)$ does not have a fibre functor over $\mathbb{Q}$, but the choice of a fibre functor over $\mathbb{Q}^{\mathrm{a}}$ identifies it with the category of representations of an affine groupoid scheme. Tannakian theory determines the "kernel" of the groupoid to be the pro-torus P and the fibre functors at all ℓ and of a polarization determine the class of the groupoid scheme in $H^2(\mathbb{Q}_l, P)$ for all l. Langlands and Rapoport (1987) showed that this family of local cohomology classes arises from a global cohomology class in $H^2(\mathbb{Q}, P)$, and hence showed that there does exist a groupoid scheme with the properties expected of that attached to $\mathsf{Mot}(k)$ and a fibre functor over $\mathbb{Q}^{\mathrm{a}}$. This is explained in my article Milne 1992.

There is a functor from affine groupoid schemes to affine extensions that can be made into an equivalence. Since extensions are more elementary than groupoid schemes, I have explained everything here in terms of them. This is a little like identifying an algebraic group over a field k with its group of k^{a}-points plus a k-structure. Purists can read the articles mentioned above.

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 realization of a groupoid. See also Milne 1992. In the above, I have ignored uniqueness questions, which can be difficult (see Milne 2003).

[84] Abelian varieties over finite fields have applications to coding theory and cryptography. Perhaps false abelian varieties, being more elementary, also have such applications.

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 $\mathbb{Q}^{\mathrm{a}}$.[85] For example, for the Siegel modular variety attached to a symplectic space (V,ψ), $\mathrm{Sh}_K(\mathbb{Q}^{\mathrm{a}})$ classifies the isomorphism classes of triples $(A,s,\eta K)$ in which A is an abelian variety defined over $\mathbb{Q}^{\mathrm{a}}$, 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 $\mathrm{Sh}_K(\mathbb{Q}^{\mathrm{a}})$ 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 $\mathbb{Q}^{\mathrm{a}}$. Suppose we know that σA is isomorphic to A for all $\sigma\in\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/E)$. 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 $H^2(\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/E),\mathrm{Aut}(A))$.

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 $L=\mathbb{Q}^{\mathrm{a}}$. 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 $\mathrm{Gal}(L^{\mathrm{a}}/L)$.

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 $\mathrm{Sh}_K(L)=\mathrm{Sh}_K(L^{\mathrm{a}})^{\mathrm{Gal}(L^{\mathrm{a}}/L)}$.

[85] This is a heuristic principle that could be made precise.

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 but not necessarily at the other primes. 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 $\mathbb{Z}_p$-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 $\Lambda\to\Lambda^\vee$. Then ψ on Λ in induces a nondegenerate alternating pairing $\Lambda/p\Lambda\times\Lambda/p\Lambda\to\mathbb{F}_p$, and $\mathcal{G}_{\mathbb{F}_p}$ is its group of symplectic similitudes.

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.

It is known (Tits 1979, 1.10) that 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 inverse system of varieties $\mathrm{Sh}_{K^p\times K_p}(G,X)$ with K^p running over the compact open subgroups of $G(\mathbb{A}_f^p)$, where $\mathbb{A}_f^p=\prod_{\ell\neq p,\infty}(\mathbb{Q}_\ell,\mathbb{Z}_\ell)$. 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.

We first state a theorem and then explain it.

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 at $\mathfrak{p}$, 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 a finite 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 often obtain another by blowing up a smooth subvariety of the closed fibre. By $\mathrm{Sh}_p(G,X)$ having *canonical* good reduction at $\mathfrak{p}$, we mean that, for any formally smooth scheme T over $\hat{\mathcal{O}}_{\mathfrak{p}}$, the natural map

$$\mathrm{Hom}_{\hat{\mathcal{O}}_{\mathfrak{p}}}(T, \varprojlim_{K^p} \mathcal{S}_{K^p}) \to \mathrm{Hom}_{E_{\mathfrak{p}}}(T_{E_{\mathfrak{p}}}, \varprojlim_{K^p} \mathrm{Sh}_{K^pK_p}) \tag{69}$$

is an isomorphism. 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 $\hat{\mathcal{O}}_{\mathfrak{p}}$, (69) characterizes it uniquely up to a unique isomorphism (by the Yoneda lemma).

(c) In the Siegel case, Theorem 16.4 was proved by Mumford (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[86] 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. See also Kisin 2010 and the later papers of Vasiu.

(d) That Sh_p should have good reduction when K_p is hyperspecial was conjectured in Langlands 1976, p. 411. That there should be a canonical model characterized by a condition like that in (b) was conjectured in Milne 1992, §2.

Definition of the Langlands-Rapoport set

Let (G,X) be a Shimura datum for which SV4,5,6 hold, and let

$$\mathrm{Sh}_p(\mathbb{C}) = \mathrm{Sh}(\mathbb{C})/K_p = \varprojlim_{K^p} \mathrm{Sh}_{K^pK_p}(G,X)(\mathbb{C}).$$

[86]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 Deligne's theory of absolute Hodge classes works is defined only for varieties over a field of characteristic zero.

For $x \in X$, let $I(x)$ be the subgroup $G(\mathbb{Q})$ fixing x, and let[87]

$$S(x) = I(x)\backslash X^p(x) \times X_p(x), \quad X^p(x) = G(\mathbb{A}_f^p), \quad X_p(x) = G(\mathbb{Q}_p)/K_p.$$

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 $G(\mathbb{Q}^{\mathrm{a}})$ such that

$$\phi'(p) = g\cdot\phi(p)\cdot g^{-1} \quad \text{for all } p \in \mathfrak{P}.$$

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 $\mathfrak{P} \to \mathrm{GL}(V(\mathbb{Q}^{\mathrm{a}})) \rtimes \Gamma$ (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,

$$I(\phi) = \{g \in G(\mathbb{Q}^{\mathrm{a}}) \mid \mathrm{ad}(g)\circ\phi = \phi\}.$$

Note that if $\rho\colon G \to \mathrm{GL}(V)$ is a faithful representation of G, then $\rho\circ\phi\colon\mathfrak{P} \to E_V$ 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 $\mathbb{Q}^{\mathrm{a}}$ dividing ℓ, and define $\mathbb{Q}_\ell^{\mathrm{a}}$ and $\Gamma_\ell \subset \Gamma$ as on p. 138.

Regard Γ_ℓ as an $\mathbb{Q}_\ell^{\mathrm{a}}/\mathbb{Q}_\ell$-affine extension with trivial kernel, and write ξ_ℓ for the homomorphism

$$\sigma \mapsto (1,\sigma)\colon \Gamma_\ell \to E_G(\ell), \quad E_G(\ell) = G(\mathbb{Q}_\ell^{\mathrm{a}}) \rtimes \Gamma_\ell.$$

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

$$X_\ell(\phi) = \mathrm{Isom}(\xi_\ell, \phi(\ell)\circ\zeta_\ell).$$

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 $G(\mathbb{Q}_\ell)$.

[87] The sets $X^p(x)$ and $X_p(x)$ do not depend on x, but it is useful to index them by x.

Note that if $\rho: G \to \mathrm{GL}(V)$ is a faithful representation of G, then

$$X_\ell(\phi) \subset \mathrm{Isom}(V(\mathbb{Q}_\ell), V_\ell(\rho \circ \phi)). \tag{70}$$

By choosing the ζ_ℓ judiciously (cf. p. 139), 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 $\mathbb{Q}^{\mathrm{a}}$ dividing p, and we use the notation of p. 139. 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 be the 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 $\rho: G \to \mathrm{GL}(V)$ is a faithful representation of G, then $D(\rho \circ \phi)$ is $V(L)$ with F acting as $v \mapsto b(\phi)\sigma v$.

Recall p. 112 that we have a well-defined $G(\mathbb{Q}^{\mathrm{a}})$-conjugacy class $c(X)$ of cocharacters of $G_{\mathbb{Q}^{\mathrm{a}}}$. We can transfer this to conjugacy class of cocharacters of $G_{\mathbb{Q}_p^{\mathrm{a}}}$, 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

$$\Phi(g) = b(\phi) \cdot \sigma b(\phi) \cdot \cdots \cdot \sigma^{m-1} b(\phi) \cdot \sigma^m g$$

where $m = [E_v : \mathbb{Q}_p]$.

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 $\phi: \mathfrak{P} \to E_G$ 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 (68) 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 $\zeta_\infty : E_\infty \to \mathfrak{P}(\infty)$ 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 $\mathfrak{P}(\infty)$. □

PROPOSITION 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 $E_\infty \to E_G(\infty)$. 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 $\phi(\infty) \circ \zeta_\infty \in \xi_X$.

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 $\phi(\ell) \circ \zeta_\ell$ 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 $\nu : G \to T$ 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 $\phi_\mu : \mathfrak{P} \to E_T$. 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 homomorphisms $\phi : \mathfrak{P} \to E_G$. 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*

$$\mathrm{LR}(G,X) \to \overline{\mathrm{Sh}}_p(G,X)(\mathbb{F}) \tag{71}$$

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[88] $\mathfrak{Q}$.

(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 p. 62) defines a map of the associated Langlands-Rapoport sets, and we should add to the conjecture that

$$\begin{array}{ccc} LR(G,X) & \longrightarrow & \overline{\mathrm{Sh}}_p(G,X)(\mathbb{F}) \\ \downarrow & & \downarrow \\ LR(T,Y) & \longrightarrow & \overline{\mathrm{Sh}}_p(T,Y)(\mathbb{F}) \end{array}$$

commutes.

(e) For the Shimura varieties defined by quaternion algebras over totally real fields, it is possible to deduce Conjecture 16.8 from the theorem of Honda and Tate (Milne 1979b,a; Reimann 1997). For more general PEL Shimura varieties, this theorem yields only weaker results.

[88]This is done in the original paper of Langlands and Rapoport, but their definition is of $\mathfrak{Q}$ is wrong. For a correct definition, see Pfau 1996.

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 (71), 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 notation 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 whose image in $G_{\mathbb{R}}^{\mathrm{ad}}$ is anisotropic),
- ◇ $\gamma=(\gamma(\ell))_{\ell\neq p,\infty}$ is an element of $G(\mathbb{A}_f^p)$ such that, for all ℓ, $\gamma(\ell)$ becomes conjugate to γ_0 in $G(\mathbb{Q}_\ell^{\mathrm{a}})$,
- ◇ δ is an element of $G(L_n)$ such that

$$\mathcal{N}\delta \overset{\mathrm{def}}{=} \delta\cdot\sigma\delta\cdot\ldots\cdot\sigma^{n-1}\delta,$$

  becomes conjugate to γ_0 in $G(\mathbb{Q}_p^{\mathrm{a}})$.

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 [why?];
- ◇ $I_\infty=$ 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 [conjugate in $G(\mathbb{Q}_\ell^a)$], there exists an isomorphism $a_\ell\colon I_{0,\mathbb{Q}_\ell^{\mathrm{a}}}\to I_{\ell,\mathbb{Q}_\ell^{\mathrm{a}}}$, well-defined up to an inner automorphism of I_0 over $\mathbb{Q}_\ell^{\mathrm{a}}$. 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 $\mathbb{Q}^{\mathrm{a}}$), 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 of $(I_0)_{\mathbb{Q}_\ell^{\mathrm{a}}}$— our assumption (*) guarantees the existence of such a system [j_ℓ and a are defined over different fields. Is $j_\ell\circ a$ defined over $\mathbb{Q}_\ell^{\mathrm{a}}$. P14]. Moreover, any other such system is isomorphic to one of the form $(I,a,(j_\ell\circ\mathrm{ad}(h_\ell)))$,[it seems to be being claimed that the I and a are the same. Why is this so?] 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 $I^{\text{ad}}(\mathbb{A})$ [Why?]. Write $d\bar{x}$ 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

$$O_\gamma(f^p) = \int_{G(\mathbb{A}_f^p)_\gamma \backslash G(\mathbb{A}_f^p)} f^p(x^{-1}\gamma x)\, d\bar{x}$$

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 $I^{\text{ad}}(\mathbb{Q}_p)$. Write $d\bar{y}$ for the quotient of dy by di_p. Proceeding as on p. 145, 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

$$TO_\delta(\varphi) = \int_{I(\mathbb{Q}_p) \backslash G(L_n)} \varphi(y^{-1}\delta\sigma(y))d\bar{y}$$

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

$$I(\gamma_0;\gamma,\delta) = I(\gamma_0;\gamma,\delta)(f^p, r) = \text{vol}(I(\mathbb{Q})\backslash I(\mathbb{A}_f)) \cdot O_\gamma(f^p) \cdot TO_\delta(\phi_r).$$

[ϕ_r has not been defined.]

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 $\phi: \mathfrak{P} \to E_G$ and a $\gamma \in I_\phi(\mathbb{Q})$ [$I_\phi(\mathbb{Q})$ has not been defined] such that $\gamma_0 x = \Phi^r x$ for some $x \in X_p(\phi)$. Here $r = [k(\mathfrak{p}){:}\mathbb{F}_p]$. An ***isomorphism*** $(\phi, \gamma_0) \to (\phi', \gamma_0')$ of admissible pairs is an isomorphism $\phi \to \phi'$ sending γ_0 to γ_0', i.e., it is a $g \in G(\mathbb{Q}^{\text{a}})$ such that

$$\text{ad}(g) \circ \phi = \phi', \quad \text{ad}(g)(\gamma_0) = \gamma_0'.$$

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 $\phi_x: \mathfrak{P} \to E_T$. 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 γ_0 belonging to a triple, the kernel of

$$H^1(\mathbb{Q}, I_0) \to H^1(\mathbb{Q}, G) \oplus \prod\nolimits_l H^1(\mathbb{Q}_l, I_0)$$

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*

$$\#\mathrm{Sh}_p(\mathbb{F}_q) = \sum_{(\gamma_0;\gamma,\delta)} c(\gamma_0) \cdot I(\gamma_0; \gamma, \delta) \tag{72}$$

where the sum is over a set of representatives for the effective triples.

PROOF. See Milne 1992, 6.13. □

NOTES. Early versions of (72) 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 (72) 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.[89] 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).

The following appendices were not in the original 2004 article.

[89] At least in the case that the weight is rational — Kottwitz does not make this assumption.

Appendix A Complements

In this section, we add some details omitted from the main text.

The notion of a moduli variety

A.1. Let k be a field. A ***moduli problem*** $(\mathcal{M},\sim)$ ***over*** k consists of a contravariant functor $\mathcal{M}$ from the category of algebraic varieties over k to the category of sets, and equivalence relations $\sim$ on each of the sets $\mathcal{M}(T)$ that are compatible with morphisms in the sense that

$$m \sim m' \implies \varphi^*(m) \sim \varphi^*(m'), \quad m,m' \in \mathcal{M}(S), \quad \varphi\colon T \to S.$$

A point t of a variety T with coordinates in k can be regarded as a map $\mathrm{Spec}(k) \to T$, and so defines a map

$$m \mapsto m_t \overset{\text{def}}{=} t^* m\colon \mathcal{M}(T) \to \mathcal{M}(k).$$

A ***solution to the moduli problem*** is a variety V over k together with a bijection

$$\alpha\colon \mathcal{M}(k)/\!\sim\; \to V(k)$$

with the properties:

(a) For any variety T over k and $m \in \mathcal{M}(T)$, the map $t \mapsto \alpha(m_t)\colon T(k) \to V(k)$ is regular (i.e., defined by a morphism of algebraic varieties);

(b) ***(Universality)*** Let Z be a variety over k and let $\beta\colon \mathcal{M}(k) \to Z(k)$ be a map such that, for any pair (T,m) as in (a), the map $t \mapsto \beta(f_t)\colon T(k) \to Z(k)$ is regular; then the map $\beta \circ \alpha^{-1}\colon V(k) \to Z(k)$ is regular.

A variety V that occurs as the solution of a moduli problem is called a ***(coarse) moduli variety***. The moduli variety (V,α) is ***fine*** if there exists a universal $m_0 \in \mathcal{M}(V)$, i.e., an object such that, for all varieties T over k and $m \in \mathcal{M}(T)$, there exists a unique regular map $\varphi\colon T \to V$ such that $\varphi^* m_0 \approx m$. Then V represents the functor $T \mapsto \mathcal{M}(T)/\sim$.

PROPOSITION A.2. *Up to a unique isomorphism, there exists at most one solution to a moduli problem.*

PROOF. Suppose there are two solutions (V,α) and (V',α'). Then because of the universality of (V,α), $\alpha' \circ \alpha^{-1}\colon V(k) \to V'(k)$ is a regular map, and because of the universality of (V',α'), its inverse is also a regular map. □

Of course, in general there may exist no solution to a moduli problem, and when there does exist a solution, it may be very difficult to prove it.

The above definitions can be stated also for the category of complex manifolds: simply replace "algebraic variety" by "complex manifold" and "regular map" by "holomorphic (or complex analytic) map". Proposition A.2 clearly also holds in this context.

The congruence subgroup problem

A.3. Let G be a semisimple group over $\mathbb{Q}$. The arithmetic and congruence subgroups of $G(\mathbb{Q})$ define topologies on it, and we denote the corresponding completions by $\hat{G}$ and $\bar{G}$.

Because every arithmetic group is congruence, the identity map on $G(\mathbb{Q})$ gives a surjective homomorphism $\hat{G} \to \bar{G}$, whose kernel $C(G)$ is called the ***congruence kernel***. This kernel is trivial if and only if all arithmetic subgroups are congruence. The modern congruence subgroup problem is to compute $C(G)$. For example, the group $C(\mathrm{SL}_2)$ is infinite.

Now let G be simply connected, and let $G' = G/N$, where N is a nontrivial subgroup of $Z(G)$. Consider the diagram:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & C(G) & \longrightarrow & \hat{G} & \longrightarrow & \bar{G} & \longrightarrow & 1 \\ & & \downarrow & & \downarrow{\scriptstyle \hat{\pi}} & & \downarrow{\scriptstyle \bar{\pi}} & & \\ 1 & \longrightarrow & C(G') & \longrightarrow & \hat{G}' & \longrightarrow & \bar{G}' & \longrightarrow & 1. \end{array}$$

It follows from the strong approximation theorem (4.16) that $\bar{G} = G(\mathbb{A}_f)$, and it follows from (3.2) that the kernel of $\hat{\pi}$ is $N(\mathbb{Q})$, which is finite. On the other hand, the kernel of $\bar{\pi}$ is $N(\mathbb{A}_f)$, which is infinite. Because $\mathrm{Ker}(\bar{\pi}) \neq N(\mathbb{Q})$, $\pi: G(\mathbb{Q}) \to G'(\mathbb{Q})$ doesn't map congruence subgroups to congruence subgroups, and because $\mathrm{Ker}(\hat{\pi}) \to \mathrm{Ker}(\bar{\pi})$ is not surjective, $C(G) \to C(G')$ is not surjective, and so $G'(\mathbb{Q})$ contains a noncongruence arithmetic subgroup. See Serre 1967b for more details.

For surveys of the congruence subgroup problem, see Raghunathan 2004 and Prasad and Rapinchuk 2008.

Proof of Theorem 5.4 (real approximation)

Recall that the theorem says that $G(\mathbb{Q})$ is dense in $G(\mathbb{R})$ if G is a connected algebraic group over $\mathbb{Q}$.

A torus T over a field k is said to be ***induced*** (or ***quasisplit***) if it is a product of tori of the form $\mathrm{Res}_{F/k}\,\mathbb{G}_m$ with F a finite separable extension of k. For such a torus T, Shapiro's lemma and Hilbert's Theorem 90 imply that $H^1(k,T) = 0$.

STEP 1. *Case that the algebraic group is a torus T over $\mathbb{Q}$.*

If T is induced, then the weak approximation theorem in algebraic number theory implies that $T(\mathbb{Q})$ is dense in $T(\mathbb{R})$.

Let T be a torus over $\mathbb{Q}$, and let F be a finite Galois extension of k splitting T with Galois group Γ. From an $x \in X_*(T)$, we get a homomorphism $[\gamma] \mapsto \gamma x : \mathbb{Z}[\Gamma/\Delta] \to X_*(T)$ where Δ is the subgroup of Γ fixing x. On applying this observation to enough elements x, we get an exact sequence

$$0 \to M_2 \to M_1 \to X_*(T) \to 0 \tag{73}$$

of Γ-modules with M_1 a finite direct sum of modules of the form $\mathbb{Z}[\Gamma/\Delta]$ (varying Δ) and $M_1^{\Delta} \to X_*(T)^{\Delta}$ surjective for all subgroups Δ of Γ. It follows from the cohomology sequence of (73) that $H^1(\Delta, M_2) = 0$ for all subgroups Δ of Γ. The sequence (73) is the cocharacter sequence of an exact sequence of tori

$$0 \to T_2 \to T_1 \to T \to 0 \tag{74}$$

with T_1 induced. The cohomology sequence of (74) is an exact sequence

$$T_1(\mathbb{R}) \to T(\mathbb{R}) \to H^1(\mathbb{C}/\mathbb{R}, T_2).$$

But

$$H^1(\mathbb{C}/\mathbb{R}, T_2) \simeq H^{-1}_{\mathrm{Tate}}(\langle\iota\rangle, M_2) \simeq H^1(\langle\iota\rangle, M_2) = 0$$

where ι denotes complex conjugation (the first isomorphism is Tate-Nakayama, and the second is the periodicity of the cohomology of cyclic groups). Therefore, $T_1(\mathbb{R})$ maps onto $T(\mathbb{R})$, and so the real approximation theorem for T follows from that for T_1.

STEP 2. *If S is an algebraic group of multiplicative type over $\mathbb{Q}$, then the map $H^1(\mathbb{Q}, S) \to H^1(\mathbb{R}, S)$ is surjective.*

The same argument as in the proof of Step 1, shows that $X^*(S)$ is a quotient of a direct sum of modules of the form $\mathbb{Z}[\Gamma/\Delta]$, and correspondingly there is an exact sequence

$$0 \to S \to T_1 \to T_2 \to 0$$

with T_1 and T_2 tori and T_1 induced. From the diagram

$$\begin{tikzcd} T_1(\mathbb{Q}) \arrow[r] \arrow[d] & T_2(\mathbb{Q}) \arrow[r] \arrow[d, "\text{dense}"] & H^1(\mathbb{Q}, S) \arrow[r] \arrow[d] & 0 \\ T_1(\mathbb{R}) \arrow[r, "\text{onto } T_2(\mathbb{R})^+"] & T_2(\mathbb{R}) \arrow[r] & H^1(\mathbb{R}, S) \arrow[r] & 0 \end{tikzcd}$$

we see that $H^1(\mathbb{Q}, S) \to H^1(\mathbb{R}, S)$ is surjective.

STEP 3. *Case that G is unirational and $G(\mathbb{R})$ is connected.*

To say that G is unirational means that there is a dominant morphism $f: U \to G$ with U a nonempty open subset of some projective space $\mathbb{P}^n$. After possibly shrinking U, we may suppose that f is smooth, and hence that $f(\mathbb{R}): U(\mathbb{R}) \to G(\mathbb{R})$ is open (for the real topology). As $\mathbb{P}^n(\mathbb{Q})$ is dense in $\mathbb{P}^n(\mathbb{R})$, the subset $U(\mathbb{Q})$ is dense in $U(\mathbb{R})$. Now $G(\mathbb{Q})$ contains $f(U(\mathbb{Q}))$, which is dense in the nonempty open subset $f(U(\mathbb{R}))$ of $G(\mathbb{R})$. The closure of $G(\mathbb{Q})$ in $G(\mathbb{R})$ is a group, and hence is also open (because its complement is a finite union of cosets). But $G(\mathbb{R})$ is connected, and so this closure equals $G(\mathbb{R})$.

STEP 4. *Case that G is a reductive group.*

Let G be a reductive group with centre Z. Then Z is of multiplicative type, and the argument in the proof of Step 1 shows that there is a surjection $T \to Z^\circ$ with T an induced torus. As $G = G^{\mathrm{der}} Z^\circ$ (almost direct product), there is an exact sequence

$$1 \to S \to G' \times T \to G \to 1$$

with G' the simply connected covering group of G^{der} and S a group of multiplicative type. The real approximation theorem holds for G' because it is unirational (Milne 2017, 17.93) and $G'(\mathbb{R})$ is connected (5.2). From the diagram

$$\begin{tikzcd} G'(\mathbb{Q}) \times T(\mathbb{Q}) \arrow[r] \arrow[d, "\text{dense image}"] & G(\mathbb{Q}) \arrow[r] \arrow[d] & H^1(\mathbb{Q}, S) \arrow[r] \arrow[d, "\text{onto}"] & H^1(\mathbb{Q}, G') \arrow[d, "\text{injective}"] \\ G'(\mathbb{R}) \times T(\mathbb{R}) \arrow[r, "\text{onto } G(\mathbb{R})^+"] & G(\mathbb{R}) \arrow[r] & H^1(\mathbb{R}, S) \arrow[r] & H^1(\mathbb{R}, G') \end{tikzcd}$$

we see that the real approximation theorem holds for G (the injectivity of the arrow at right is the Hasse principle for G', Platonov and Rapinchuk 1994, Theorem 6.6, p. 286).

STEP 5. *The general case.*

There is an exact sequence $1 \to U \to G \to G/U \to 1$ with U a normal connected unipotent algebraic subgroup of G and G/U reductive. A theorem of Mostow says that the sequence splits, and every connected unipotent group over $\mathbb{Q}$ is isomorphic (as a scheme) to affine n-space for some n. From this the statement follows. (See Milne 2017, 19.11, 25.49, and 14.32, for the statements used.)

See also Platonov and Rapinchuk 1994, Theorem 7.7, p. 415.

SECOND PROOF (FROM A LETTER OF G. PRASAD, SEPT. 1, 1987).

"To prove that $G(\mathbb{Q})$ is dense in $G(\mathbb{R})$, what you need is a result of H. Matsumoto, which is reproved in Borel-Tits "Groupes réductifs", Publ. Math. IHES no. 27, as Théorème 14.4, according to which given a maximal $\mathbb{R}$-split torus S of G, $S(\mathbb{R})$ meets every connected component of $G(\mathbb{R})$. Now we observe that there is a maximal torus T defined over $\mathbb{Q}$ which contains a maximal $\mathbb{R}$-split torus of G: To prove this, we will make use of the fact that the closure of $G(\mathbb{Q})$ contains $G(\mathbb{R})^+$. Take a maximal torus $\mathcal{T}$ defined over $\mathbb{R}$ and containing a maximal $\mathbb{R}$-split torus of G. In $\mathcal{T}(\mathbb{R})^+$, let $\mathcal{U}$ be the set of *regular* elements. $\mathcal{U}$ is open in $\mathcal{T}(\mathbb{R})$. Now let $U = \bigcup_{g \in G(\mathbb{R})} g\mathcal{U}g^{-1}$; then U is an open subset of $G(\mathbb{R})^+$ (to see this, consider the map $G(\mathbb{R}) \times \mathcal{U} \to G(\mathbb{R})$ defined by $(g,x) \mapsto gxg^{-1}$; it is everywhere regular). Hence, there exists $t \in U \cap G(\mathbb{Q})$. As t is regular, the identity component of the centralizer of t in G is a torus T defined over $\mathbb{Q}$, and as t has a conjugate in $\mathcal{U}$, it is obvious that T contains a conjugate of the maximal $\mathbb{R}$-split torus in $\mathcal{T}$. This proves that there is a maximal torus defined over $\mathbb{Q}$ which contains a maximal $\mathbb{R}$-split torus of G."

Proof of the claim in 5.23

PROPOSITION A.4. *Let (G,X) be a Shimura datum with G^{der} simply connected, and assume that $Z' \overset{\mathrm{def}}{=} Z \cap G^{\mathrm{der}}$ satisfies the Hasse principle for H^1, i.e.,*

$$H^1(Q, Z') \to \prod_{l=2,3,\ldots,\infty} H^1(Q_l, Z')$$

is injective. Then, for any sufficiently small compact open subgroup K of $G(\mathbb{A}_f)$,

$$G(\mathbb{Q})_+ \cap K \subset Z(\mathbb{Q}) \cdot G^{\mathrm{der}}(\mathbb{Q}).$$

PROOF. (Cf. the proof of 5.20.) Consider the diagram:

$$\begin{array}{ccccccccc}
1 & \longrightarrow & Z'(\mathbb{Q}) & \longrightarrow & Z(\mathbb{Q}) \times G^{\mathrm{der}}(\mathbb{Q}) & \longrightarrow & G(\mathbb{Q}) & \longrightarrow & H^1(\mathbb{Q}, Z') \\
 & & \| & & \downarrow & & \downarrow & & \| \\
1 & \longrightarrow & Z'(\mathbb{Q}) & \longrightarrow & G^{\mathrm{der}}(\mathbb{Q}) & \longrightarrow & G^{\mathrm{ad}}(\mathbb{Q}) & \longrightarrow & H^1(\mathbb{Q}, Z').
\end{array}$$

Let $q \in G(\mathbb{Q})_+$. By definition, the image of q in $G^{\mathrm{ad}}(\mathbb{R})$ lies in its identity component, and so lifts to an element of $G^{\mathrm{der}}(\mathbb{R})$. Therefore, the image of q in $H^1(\mathbb{R}, Z')$ is zero. The isogeny $Z \times G^{\mathrm{der}} \to G$ extends to an étale isogeny over $\operatorname{Spec} \mathbb{Z}[d^{-1}]$ for some d. For any ℓ not dividing d, the map $Z(\mathbb{Z}_\ell) \times G^{\mathrm{der}}(\mathbb{Z}_\ell) \to G(\mathbb{Z}_\ell)$ is surjective, and so, if $q \in G(\mathbb{Z}_\ell)$, then it maps to zero in $H^1(\mathbb{Q}_\ell, Z')$. For the remaining ℓ, the map $Z(\mathbb{Z}_\ell) \times G^{\mathrm{der}}(\mathbb{Z}_\ell) \to G(\mathbb{Z}_\ell)$

will have open image K_ℓ. Therefore, if $q \in \prod_{\ell \nmid d} G(\mathbb{Z}_\ell) \times \prod_{\ell | d} K_\ell$, then it maps to zero in $\prod_{l \text{ finite}} H^1(\mathbb{Q}_l, Z')$. Because of the Hasse principle, this implies that g maps to zero in $H^1(\mathbb{Q}, Z')$, and therefore lies in $Z(\mathbb{Q}) \cdot G^{\mathrm{der}}(\mathbb{Q})$. □

Hermitian forms (details for §8)

By a k-algebra we mean a ring containing k in its centre and of finite dimension as a k-vector space. A k-algebra A is said to be ***separable*** if A is semisimple and its centre is an étale k-algebra. For such a k-algebra, the pairing

$$(a,b) \mapsto \mathrm{Tr}_{A/k}(ab) \colon A \times A \to k$$

given by the *reduced* trace $\mathrm{Tr}_{A/k}$ is nondegenerate (Curtis and Reiner 1981, 7.41).

LEMMA A.5. *Let F be an étale k-algebra and V a free F-module of finite rank. The map*

$$f \mapsto \mathrm{Tr}_{F/k} \circ f \colon \mathrm{Hom}_F(V, F) \to \mathrm{Hom}_k(V, k)$$

is an isomorphism of k-vector spaces.

PROOF. The map $f \mapsto \mathrm{Tr}_{F/k} \circ f$ is injective because the trace pairing is nondegenerate, and it is surjective because the two spaces have the same dimension. □

LEMMA A.6. *Let $(A, *)$ be a separable k-algebra with involution, and let F be its centre. Let V and W be A-modules that are free of finite rank over F, and let $\psi \colon V \times W \to k$ be an A-balanced k-bilinear form. Then there exists a unique F-sesquilinear form $\phi \colon V \times W \to F$ such that*

$$\psi(v, w) = \mathrm{Tr}_{F/k}(\phi(v, w)) \text{ for all } v \in V \text{ and } w \in W;$$

moreover, ϕ is A-balanced.

PROOF. Make V into a right A-module by letting a act as a^*. Then ψ factors through a k-linear homomorphism $\psi' \colon V \otimes_A W \to k$. Let $c \in F$ act on $V \otimes_A W$ by $c(v \otimes w) = cv \otimes w = v \otimes c^* w$. According to Lemma A.5, there is a unique F-linear map $\phi' \colon V \otimes_A W \to F$ such that $\mathrm{Tr}_{F/k} \circ \phi' = \psi'$. Now $\phi \overset{\text{def}}{=} ((v, w) \mapsto \phi'(v \otimes w))$ is an A-balanced F-sesquilinear form such that $\psi = \mathrm{Tr}_{F/k} \circ \phi$:

$$\begin{tikzcd}
V \times W \arrow[r, "{v,w \mapsto v \otimes w}"] \arrow[rr, bend left, "\phi"] \arrow[dr, "\psi"'] & V \otimes_A W \arrow[r, "\phi'"] \arrow[d, "\psi'"] & F \arrow[dl, "\mathrm{Tr}_{F/k}"] \\
& k. &
\end{tikzcd}$$

It remains to prove the uniqueness. Let ϕ_1 and ϕ_2 be F-sesquilinear forms $V \times W \to F$ such that $\mathrm{Tr}_{F/k} \circ \phi_1 = \mathrm{Tr}_{F/k} \circ \phi_2$. Let $v \in V$ and $w \in W$. Then $\mathrm{Tr}_{F/k} \phi_1(cv, w) = \mathrm{Tr}_{F/k} \phi_2(cv, w)$ for all $c \in F$, and so $\mathrm{Tr}_{F/k} c\phi_1(v, w) = \mathrm{Tr}_{F/k} c\phi_2(v, w)$ for all $c \in F$. Because of the nondegeneracy of the trace pairing, this implies that $\phi_1(v, w) = \phi_2(v, w)$. □

LEMMA A.7. *Let $(A,*)$ be a separable k-algebra with involution, and let F be its centre. Let $f \in F^{\times}$ be such that $f^* = -f$, and let (V,ψ) be a symplectic (A,ψ)-module with V free of finite rank over F. Then there exists a unique hermitian form $\phi\colon V \times V \to F$ such that*

$$\psi(u,v) = \mathrm{Tr}_{F/k}(f\phi(u,v))$$

for all $u,v \in V$; moreover ϕ is A-balanced.

PROOF. Lemma A.6 with $V = W$ gives an A-balanced F-sesquilinear form $\phi_1\colon V \times V \to F$ such that $\psi(u,v) = \mathrm{Tr}_{F/k}\,\phi_1(u,v)$. Let $\phi = f^{-1}\phi_1$, so that $\psi(u,v) = \mathrm{Tr}_{F/k}(f\phi(u,v))$. Now ϕ is sesquilinear and A-balanced, and to show that it is hermitian it remains to show that $\phi(u,v) = \phi(v,u)^*$. By assumption $\psi(u,v) = -\psi(v,u)$, and so

$$\mathrm{Tr}_{F/k}(f\phi(u,v)) = -\mathrm{Tr}_{F/k}(f\phi(v,u)) = \mathrm{Tr}_{F/k}(f^*\phi(v,u)).$$

On replacing u by cu with $c \in F$, we find that

$$\mathrm{Tr}_{F/k}(fc\phi(u,v)) = \mathrm{Tr}_{F/k}((fc)^*\phi(v,u)).$$

But $\mathrm{Tr}_{F/k}(a) = \mathrm{Tr}_{F/k}(a^*)$ for all $a \in F$, and so

$$\mathrm{Tr}_{F/k}((fc)^*\phi(v,u)) = \mathrm{Tr}_{F/k}(fc\phi(v,u)^*).$$

As fc is an arbitrary element of F, the non-degeneracy of the trace form implies that $\phi(u,v) = \phi(v,u)^*$. The uniqueness of ϕ follows from Lemma A.6. □

The $*$-action on the Dynkin diagram

A.8. Let G be a semisimple group G over $\mathbb{Q}$ and let T be a maximal torus in G. Then $(G,T)_{\mathbb{Q}^{\mathrm{a}}}$ is split and so the choice of a Borel subgroup B of $G_{\mathbb{Q}^{\mathrm{a}}}$ containing $T_{\mathbb{Q}^{\mathrm{a}}}$ determines a based semisimple root datum (X,Φ,Δ). There is a natural action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ on $X \overset{\text{def}}{=} X^*(T)$ which preserves the subset Φ (but not necessarily Δ because it depends on the choice of B). The Weyl group of (G,T) acts simply transitively on the set of bases for Φ. If $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$, then $\sigma(\Delta)$ is also a base for Φ, and so $w_\sigma(\sigma(\Delta)) = \Delta$ for a unique $w_\sigma \in W$. For $\alpha \in \Delta$, define $\sigma * \alpha = w_\sigma(\sigma\alpha)$. One checks that this defines an action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{a}}/\mathbb{Q})$ on Δ. When G is split, we can choose T to be a split maximal torus in G and B to be a Borel subgroup of G (i.e., defined over $\mathbb{Q}$) containing T. In this case, the action is trivial. As G splits over a finite extension of $\mathbb{Q}$, this shows that the action is continuous. When G is quasi-split, we can choose a Borel pair (B,T) in G and define (X,Φ,Δ) to be the based semisimple root datum attached to $(G,B,T)_{k^{\mathrm{s}}}$. In this case, Δ is stable under the action of Γ on X, and the $*$-action on Δ is induced by the natural action of Γ on X. See Milne 2017, 24.6.

Appendix B List of Shimura varieties of abelian type

We give a list of classical reductive groups G_0 such that (G_0, X_0) defines a Shimura variety for a suitable X_0, and such that $(G_0^{\mathrm{der}}, X_0)$ is of primitive abelian type. Every (G, X^+) of primitive abelian type is of the form $(G_0^{\mathrm{der}}, X_0)$ with (G_0, X_0) from the following list. These (G_0, X_0) all have the property that $E(G_0, X_0) = E(G_0^{\mathrm{ad}}, X_0^+)$.

In the following, F_0 is a totally real number field, and I is the set of all embeddings of F_0 into $\mathbb{R}$. We use $\bar{z}$ to denote the complex conjugate of a complex number z.

(A) Let K be a quadratic totally imaginary extension of F_0 and A a central simple algebra over K equipped with an involution σ of the second kind. There is a reductive group G' over F_0 such that

$$G'(F_0) = \{x \in A^\times \mid xx^\sigma \in F_0^\times\}.$$

We put $G_0 = (G')_{F_0/\mathbb{Q}}$. The center of G_0 is $(\mathbb{G}_m)_{K/\mathbb{Q}}$.

For non-negative integers r and s, we put

$$I_{r,s} = \begin{pmatrix} I_r & 0 \\ 0 & -I_r \end{pmatrix}$$

and

$$GU(r,s) = \{g \in \mathrm{GL}_{r+s}(\mathbb{C}) \mid gI_{r,s}\bar{g}^t = \nu(g)I_{r,s},\ \nu(g) \in \mathbb{R}^\times\}.$$

Then, for each $v \in I$, there are non-negative integers r_v and s_v such that

$$G_0(\mathbb{R}) \approx \prod_v GU(r_v, s_v).$$

Let $I_{nc} = \{v \in I \mid r_v \cdot s_v \neq 0\}$ and let I_c be the complement of I_{nc}. Define $h_v{:}\,\mathbb{S} \simeq \mathbb{C}^\times \to GU(r_v, s_v)$ by

$$h_v(z) = \begin{cases} \begin{pmatrix} zI_{r_v} & 0 \\ 0 & \bar{z}I_{s_b} \end{pmatrix} & \text{if } v \in I_{nc} \\ 1 & \text{if } v \in I_c, \end{cases}$$

and define $h_0{:}\,\mathbb{S} \to G_0(\mathbb{R})$ to be the product of the h_v. Let X_0 be the $G_0(\mathbb{R})$-conjugacy class of h_0. Then (G_0, X_0) defines a Shimura variety. For any connected component X_0^+ of X_0, the pair $(G_0^{\mathrm{der}}, X_0^+)$ is of type A.

The reflex field $E(G_0, X_0)$ is either $\mathbb{Q}$ or a CM-field. The first case happens if and only if $r_v = s_v$ for all $v \in I$. In this case the map η defined in Section 3 takes h_0 to $h_0' = \prod_v h_v'$, where

$$h_v(z) = \begin{cases} \begin{pmatrix} \bar{z}I_r & 0 \\ 0 & zI_r \end{pmatrix}, \quad r = r_v = s_v, & \text{if } v \in I_{nc} \\ 1 & \text{if } v \in I_c. \end{cases}$$

(B) Let $n \geq 3$ be an odd integer and q a quadratic form on an n-dimensional vector space over F_0 such that the signature of q at a $v \in I$ is $(n,0)$, $(0,n)$, $(n-2,2)$, or $(2,n-2)$. The special Clifford group of q defines a reductive group G' over F_0. We put $G_0 = (G')_{F_0/\mathbb{Q}}$. The center of G_0 is $(\mathbb{G}_m)_{F_0/\mathbb{Q}}$.

We assume that q_v has signature $(n-2,2)$ for $v = 1, \ldots, r$ $(1 \leq r \leq g)$, and is positive definite for $v > r$.We have

$$G_{\mathbb{R}} = \prod_{v=1}^{g} \mathrm{Gpn}(q_v).$$

Therefore, to define $h_o\colon\mathbb{S}\to G_{\mathbb{R}}$,we only have to define its components $\mathbb{S}\to Gpn(q_v)$.

For $v>r$, we let h_v be the trivial homomorphism. To define h_v for $v\le r$, take the even Clifford algebra E_v of q_v.Denote the main involution of $E_v$by ι. Fix two orthogonal vectors e_1, e_2 of V_v such that $q_v(e_1)=q_v(e_2)=-1$. Then $j_v=e_1e_2\in E_v$. We have $j_v^2=-1$ and $j_v^\iota=-j_v$. Furthermore, for $(a,b)\in\mathbb{R}^2\smallsetminus(0,0)$, $a+bj_v\in Gpn(q_v)$. We define $h_v\colon\mathbb{C}^\times\to Gpn(q_v)(\mathbb{C})$ by

$$h_v(a+bi)=a+bj_v.$$

Let K_∞ be the centralizer of h_0 in $G(\mathbb{R})$. Then $K_\infty\cap G'(\mathbb{R})$ is a maximal compact subgroup of $G'(\mathbb{R})$ and the quotient $G'(\mathbb{R})/K_\infty\cap G'(\mathbb{R})$ is a bounded symmetric domain isomorphic to the product of r copies of

$$\left\{(\zeta_1,\ldots,\zeta_p)\in\mathbb{C}^p\mid\sum_{k=1}^p|\zeta_k|^2<\frac{1}{2}\left(1+\left|\sum_{k=1}^p\zeta_k\right|^2\right)<1\right\}$$

where we put $p=n-2$. The field $E(G_0,X_0)$ is generated over $\mathbb{Q}$ by $\left\{\sum_{v=1}^r x^{\tau_v}\mid x\in F\right\}$

We refer to Shih 1978 for the description of X_0 such that (G_0,X_0) is a Shimura datum. The reflex field $E(G_0,X_0)$ is totally real. The derived group G_0^{der} is the spin group of q. For any connected component X_0^+ of X_0, the pair $(G_0^{\mathrm{der}},X_0^+)$ is of type B.

(C) G_0 is the similitude group of a hermitian form over a quaternion algebra whose centre is F_0; see section 7.

($\mathrm{D}^{\mathbb{R}}$) There are two cases:

(a) Same as type B , except $n\ge4$ is even.

(b) Let B be a totally indefinite quaternion algebra over F_0 and denote by σ the main involution of B. Let q be a σ-antihermitian form on a left free B-module of rank $n\ge2$. At each $\tau\in I$, q defines a quadratic form on a $2n$-dimensional real vector space. We assume that its signature is $(2n,0)$, $(0,2n)$, $(2n-2,2)$, or $(2,2n-2)$. Let G' be the algebraic group over F_0 defined by the special Clifford group of q, and let $G_0=(G')_{F_0/\mathbb{Q}}$. We define X_0 as before(?) Shih 1978. Then (G_0,X_0) is a Shimura datum of type $\mathrm{D}^{\mathbb{R}}$

In both cases $E(G_0,X_0)$ is totally real, and the center of G_0 is $(Z')_{F_0/\mathbb{Q}}$, where Z' is an extension of μ_2 by $\mathbb{G}_m$ over F_0.

($\mathrm{D}^{\mathbb{H}}$) Let B be a quaternion algebra over F_0 with main involution σ. Let q be a σ-anti-hermitian form on a free left B-module Λ of rank $n\ge4$. Let I_{nc} be the set of $\tau\in I$ where B does not split, and let I_c be the complement of I_{nc} As usual, we assume that I_{nc} is non-empty; let r be its cardinality. We assume also that at every $\tau\in I_c$, the real quadratic form defined by q is definite. Let G_0 be the algebraic group over $\mathbb{Q}$ such that

$$G_0(\mathbb{Q})=\{g\in\mathrm{GL}_B(\Lambda)\mid gq^tg^\sigma=\nu(g)q,\,\nu(g)\in F_0^\times\text{ and }N(g)=\nu(g)^n\},$$

where N denotes the reduced norm from $\mathrm{End}_B(\Lambda)$ to F_0. Then $G_0(\mathbb{R})$ is isomorphic to the product of r copies of $GO^*(2n)$, where $GO^*(2n)$ consists of the $g=\begin{pmatrix}A&B\\-\bar{B}&\bar{A}\end{pmatrix}\in$ $\mathrm{GL}_{2n}(\mathbb{C})$ such that

$$g\begin{pmatrix}I_n&0\\0&-I_n\end{pmatrix}\bar{g}^t=\nu(g)\begin{pmatrix}I_n&0\\0&-I_n\end{pmatrix},\ \nu(g)\in\mathbb{R}^\times,\ \det(g)=\nu(g)g^n.$$

Define $h_0 : \mathbb{S} \simeq \mathbb{C}^\times \to G_0(\mathbb{R}) \simeq GO^*(2n)^r$ so that each component of h_0 is given by

$$z \mapsto \begin{pmatrix} zI_n & 0 \\ 0 & \bar{z}I_n \end{pmatrix}$$

and define X_0 to be the $G_0(\mathbb{R})$-conjugacy class of h_0 Then (G_0, X_0) is a Shimura datum. The center of G_0 is $(\mathbb{G}_m)_{F_0/\mathbb{Q}}$.

The reflex field $E(G_0, X_0)$ is a CM-field or a totally real field according as n is odd or even. Let $h_0' : \mathbb{S} \simeq \mathbb{C}^\times \to G_0(\mathbb{R}) \simeq (GO^*(2n))^r$ be a homomorphism such that each component of h_0' is given by

$$z \mapsto \begin{pmatrix} \bar{z}I_n & 0 \\ 0 & zI_n \end{pmatrix}.$$

Then h_0' belongs to X_0 if and only if n is even. In this case the map η defined in 7.3 takes h_0 to h_0'.

When $n = 4$, we also allow G_0 of the "mixed type". We let I_c be the set of $\tau \in I$ such that B splits at τ and the quadratic form over $\mathbb{R}$ determined by q at τ is definite. Denote the complement of I_c by I_{nc} Let s (resp. r) be the number of $\tau \in I_{nc}$ at which B splits (resp. does not split). We assume that $r > 0$. If B splits at a $\tau \in I_{nc}$, we assume that the signature of the real quadratic form determined by q at τ is $(6,2)$ or $(2,6)$. Then

$$G_0(\mathbb{R}) \simeq GO^*(8)^r \times \left(GO(6,2)^+\right)^s,$$

where

$$GO(6,2)^s = \left\{ g \in \mathrm{GL}_8(\mathbb{R}) \,\middle|\, g \begin{pmatrix} I_6 & 0 \\ 0 & I_2 \end{pmatrix} g^t = \nu(g) \begin{pmatrix} I_6 & 0 \\ 0 & I_2 \end{pmatrix}, \nu(g) \in \mathbb{R}^\times, \det(g) > 0 \right\}.$$

We define $h_0 : \mathbb{S} \to G_0(\mathbb{R})$ to be the homomorphism whose projection to the component $GO^*(8)$ is defined as before and the projection to the component $GO(6,2)^+$ is

$$z \mapsto \begin{pmatrix} |z|^2 I_6 & 0 & 0 \\ 0 & \Re(z^2) & \Im(z^2) \\ 0 & -\Im(z^2) & \Re(z^2) \end{pmatrix}.$$

Let X_0 be the $G_0(\mathbb{R})$-conjugacy class of h_0. Then (G_0, X_0) is a Shimura datum.

The reflex field $E(G_0, X_0)$ is totally real. Let h_0' be the image of h_0 under the map η of Section 3. Then the projection of h_0' to $GO^*(8)$ is

$$z \mapsto \begin{pmatrix} \bar{z}I_4 & 0 \\ 0 & zI_4 \end{pmatrix},$$

and the projection to $GO(6,2)^+$ is

$$z \mapsto \begin{pmatrix} |z|^2 I_6 & 0 & 0 \\ 0 & \Re(z^2) & -\Im(z^2) \\ 0 & \Im(z^2) & \Re(z^2) \end{pmatrix}.$$

NOTES. This appendix is copied from the appendix of Milne and Shih 1981a.

Appendix C Review of Shimura's Collected Papers [90]

When Weil arrived in Tokyo in 1955, planning to speak about his ideas on the extension to abelian varieties of the classical theory of complex multiplication, he was surprised to learn that two young Japanese mathematicians had also made decisive progress on this topic.[91] They were Shimura and Taniyama. While Weil wrote nothing on complex multiplication except for the report on his talk, Shimura and Taniyama published their results in a book in Japanese, which, after the premature death of Taniyama, was revised and published in English by Shimura. For a polarized abelian variety with many complex multiplications, the theory describes the action of the absolute Galois group of a certain reflex field on the moduli of the variety and its points of finite order, and it expresses the zeta function of the abelian variety in terms of Hecke L-series. Over the years Shimura found various improvements to these original results, which are included among the papers collected in these volumes.

Complex multiplication is the foundation stone of what has become known as the theory of canonical models. Each elliptic modular curve is defined in a natural way over a number field k (which depends on the curve). For analysts, the explanation for this is that the Fourier expansions at the cusps provide a k-structure on the spaces of modular functions and forms. For geometers, the explanation is that the curve is the solution of a moduli problem which is defined over k. In one of his most significant results, Shimura showed that quotients of the complex upper half plane by quaternionic congruence groups are also naturally defined over number fields, even when compact (hence without cusps) and even when they are not moduli varieties in any natural way.[92] As the fruit of a long series of investigations, he found a precise notion of a canonical model for the congruence quotients of bounded symmetric domains and proved that they exist for important families. Let G be a semisimple algebraic group over $\mathbb{Q}$ such the quotient of $G(\mathbb{R})$ by a maximal compact subgroup is a bounded symmetric domain X. Then each quotient $\Gamma\backslash X$ of X by a congruence subgroup Γ of $G(\mathbb{Q})$ has a model over a specific number field k_Γ. As Γ varies, these models are compatible, and the whole family is called the canonical model. It is characterized by reciprocity laws at the CM-points, and its definition requires a realization of G as the derived group of a reductive group.[93]

In his talk at the International Congress in 1978,[94] Shimura raised five questions: I. Can one define the notion of arithmetic automorphic functions? II. Same question for automorphic forms. III. Are (holomorphic) Eisenstein series arithmetic? IV. Is there any explicit way to construct arithmetic automorphic forms, similar to Eisenstein series, in the case of compact quotient? V. Is there any interpretation of the values of such explicit arithmetic automorphic forms at CM-points as [critical] values of zeta functions? Question I is answered by the theory of canonical models: the model of $\Gamma\backslash X$ over k_Γ provides a k_Γ-structure on the space of automorphic functions for Γ. Question II asks whether there is (at least) a natural $\bar{\mathbb{Q}}$-structure on the holomorphic automorphic forms. Much of Shimura's work over the last

[90] This is MR1965574. The footnotes, which date from 2003, are not in the review sent to MR.

[91] Weil, CW II, p. 541. He was in Tokyo for the famous International Symposium on Algebraic Number Theory at Tokyo and Nikko. See my webpage under documents.

[92] This work is described in his article NAMS 43 (1996), no. 11, 1340–1347 (CW IV, p. 491).

[93] See Shimura's talk at the ICM 1970 (Nice) (CW II, p. 400) for more precise statements and a statement of the "Shimura conjecture".

[94] CW III, p. 147.

twenty-five years has been directed towards answering these questions, especially question V. This has involved the study of the periods of abelian varieties, Eisenstein series, differential operators on bounded symmetric domains, and the notion of near holomorphy. (For a more extensive overview of Shimura's work, I recommend the article of H. Yoshida, Bull. AMS, 39 (2002), 441-448.)

The four volumes under review collect all the papers published by Shimura between 1954 and 2001 except for a few which are mainly expository. It also includes three articles not previously published[95] and two articles published only in mimeographed proceedings of the conferences, and hence not generally available.[96] Most papers are reproduced directly from the originals, but fifteen have been newly typeset (not without the introduction of new misprints), including three[97] that were re-typeset from the author's manuscripts because of errors introduced into the published versions by incompetent typesetters and copyeditors.

Over fifty pages of endnotes have been added. Most notes correct misprints or other minor errors, but some give more extended clarifications or complements to the papers. The origins of the conjecture on the modularity of elliptic curves are revisited in the endnotes to the papers [64e] and [89a] (and also in the article [96b] itself).

In the preface, Shimura writes: "Some of my recollections are included with the hope that they may help the reader have a better perspective. I have also mentioned the results in my later articles which supersede or are related to those in the article at issue. However, I decided not to mention the results of other later investigators, mainly in order to make my task easier". Thus, the endnotes help place the individual papers in the general context of Shimura's work,[98] but not in any wider context. In fact, the work whose origins can be traced to that of Shimura is extensive.

Whereas reductive groups play a somewhat auxiliary role in Shimura's work, Deligne adopted them as the starting point in his 1969[99] Bourbaki report on Shimura's work. There is now a large body of work on what are called Shimura varieties, expressed in the language of abstract reductive groups (roots and weights) and Grothendieck algebraic geometry (schemes and motives). In this more general context, the existence of canonical models had been proved for all Shimura varieties, including those attached to the exceptional groups, by 1982,[100] and by 1986 the theory of automorphic vector bundles had yielded in complete generality a notion of the arithmeticity of holomorphic automorphic forms over the reflex field (or even $\mathbb{Q}$).[101] Moreover, by 1982 the main theorems of complex multiplication had been extended to all automorphisms of $\bar{\mathbb{Q}}$ (not just those fixing the reflex field).[102] Thus, by the mid 1980s, it was possible to ask some of the arithmeticity questions mentioned above over $\mathbb{Q}$.

[This paragraph was my attempt to briefly place Shimura's work in context. In fact, since about 1970 there have been two schools in the field, which I'll refer to as the Shimura school and the Deligne school. In terms of the number of published papers, the first is much larger than the second (this was written in 2003).

[95] 1968c, 2001b, 2001c.

[96] 1963e, 1964e.

[97] 1967c, 1978c, 1997b.

[98] In his papers (including his survey papers), his books, and in the comments in his Collected Papers, Shimura ignores almost all work not done by himself or his students. Consequently, a mathematician studying only his writings will get an incorrect impression of what is known in the field.

[99] Actually, the report was in February 1971

[100] Deligne, Borovoi, Milne.

[101] Brylinski, Harris, Milne.

[102] Deligne, Langlands

I will describe what I see as the essential difference between the two schools. Initially one begins with a semisimple group G over $\mathbb{Q}$ with $G(\mathbb{R})$ acting on a hermitian symmetric domain D (satisfying certain conditions). As Shimura first understood, to get a canonical model one needs to realize G as the derived group of a reductive group. For Shimura the reductive group is auxiliary: given G he makes the most convenient choice for the reductive group. On the other hand, Deligne begins with the reductive group. Different choices of the reductive group for a given G give different canonical models, but they all give the same connected canonical model (in the sense of Deligne — it is an inverse system of connected varieties over $\bar{\mathbb{Q}}$ with an action of a big group). I think that for most of what he does, Shimura only needs the connected canonical model (and, in general, that's all his theory gives). Thus, except for special Shimura varieties (those given by his choice of the reductive group), his is intrinsically a $\bar{\mathbb{Q}}$-theory, whereas Deligne's is a $\mathbb{Q}$-theory. The challenge is to rewrite all of the work done by the Shimura school in Deligne's language. This will mean, for example, that when the Shimura school proves that some special value is ap where a is an algebraic number and p is a transcendental period, one should prove that a lies in an abelian extension of a specific number field and describe how the Galois group acts (and, when the Shimura school obtains the finer result for special Shimura varieties, one should obtain it for general Shimura varieties).]

References

ASH, A., MUMFORD, D., RAPOPORT, M., AND TAI, Y. 1975. Smooth compactification of locally symmetric varieties. Math. Sci. Press, Brookline, Mass.

BAILY, W. L. 1958. Satake's compactification of V_n. *Amer. J. Math.* 80:348–364.

BAILY, JR., W. L. AND BOREL, A. 1966. Compactification of arithmetic quotients of bounded symmetric domains. *Ann. of Math. (2)* 84:442–528.

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, pp. 23–40. *In* Colloq. Théorie des Groupes Algébriques (Bruxelles, 1962). Librairie Universitaire, Louvain.

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. Springer-Verlag, New York.

BOREL, A. 1998. Semisimple groups and Riemannian symmetric spaces, volume 16 of *Texts and Readings in Mathematics*. Hindustan Book Agency, New Delhi.

BOREL, A. 2001. Essays in the history of Lie groups and algebraic groups, volume 21 of *History of Mathematics*. American Mathematical Society, Providence, RI.

BOREL, A. AND HARISH-CHANDRA 1962. Arithmetic subgroups of algebraic groups. *Ann. of Math. (2)* 75:485–535.

BOREL, A. AND JI, L. 2006. Compactifications of symmetric and locally symmetric spaces. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA.

BOURBAKI, N. 1958. 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.

BRYLINSKI, J.-L. 1983. "1-motifs" et formes automorphes (théorie arithmétique des domaines de Siegel), pp. 43–106. *In* Conference on automorphic theory (Dijon, 1981), volume 15 of *Publ. Math. Univ. Paris VII*. Univ. Paris VII, Paris.

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.

CASSELMAN, W. A. 1997. Geometric rationality of Satake compactifications, pp. 81–103. *In* Algebraic groups and Lie groups, Austral. Math. Soc. Lect. Ser. Cambridge Univ. Press, Cambridge.

CHOW, W.-L. 1949. On compact complex analytic varieties. *Amer. J. Math.* 71:893–914.

CURTIS, C. W. AND REINER, I. 1981. Methods of representation theory. With applications to finite groups and orders. Vol. I. John Wiley & Sons, Inc., New York.

DELIGNE, P. 1970. Travaux de Griffiths. Séminaire Bourbaki, 22ème année (1969/70), Exp. No. 376.

DELIGNE, P. 1971a. Théorie de Hodge. II. *Inst. Hautes Études Sci. Publ. Math.* 40:5–57.

DELIGNE, P. 1971b. Travaux de Shimura. Séminaire Bourbaki, 23ème année (1970/71), Exp. No. 389.

DELIGNE, P. 1971c. Travaux de Shimura. Preliminary unpublished version of Deligne 1971b.

DELIGNE, P. 1973a. Espaces hermitiens symétriques, spring 1973. Handwritten notes.

DELIGNE, P. 1973b. Les constantes des équations fonctionnelles des fonctions L, pp. 501–597. Lecture Notes in Math., Vol. 349. *In* Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972). Springer, Berlin.

DELIGNE, P. 1979. Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques, pp. 247–289. *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.

DELIGNE, P. 1982. Hodge cycles on abelian varieties (notes by J.S. Milne), pp. 9–100. *In* Hodge cycles, motives, and Shimura varieties, Lecture Notes in Mathematics. Springer-Verlag, Berlin. Version with endnotes available at jmilne.org.

DEMAZURE, M. 1972. Lectures on p-divisible groups. Springer-Verlag, Berlin.

GIRAUD, J. 1968. Remarque sur une formule de Shimura-Taniyama. *Invent. Math.* 5:231–236.

GORESKY, M. 2005. Compactifications and cohomology of modular varieties, pp. 551–582. *In* Harmonic analysis, the trace formula, and Shimura varieties, volume 4 of *Clay Math. Proc.* Amer. Math. Soc., Providence, RI.

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 online.

HELGASON, S. 1978. Differential geometry, Lie groups, and symmetric spaces, volume 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.

HSU, T. 1996. Identifying congruence subgroups of the modular group. *Proc. Amer. Math. Soc.* 124:1351–1359.

HUMPHREYS, J. E. 1978. Introduction to Lie algebras and representation theory, volume 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, volume 76 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.

KISIN, M. 2010. Integral models for Shimura varieties of abelian type. *J. Amer. Math. Soc.* 23:967–1012.

KLEIN, F. 1880. Zur Theorie der elliptischen Modulfunktionen. *Math. Ann.* 17:62–70.

KNUS, M.-A., MERKURJEV, A., ROST, M., AND TIGNOL, J.-P. 1998. The book of involutions, volume 44 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI.

KOLLÁR, J. 2007. Lectures on resolution of singularities, volume 166 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ.

KOTTWITZ, R. E. 1984. Shimura varieties and twisted orbital integrals. *Math. Ann.* 269:287–300.

KOTTWITZ, R. E. 1990. Shimura varieties and λ-adic representations, pp. 161–209. *In* Automorphic forms, Shimura varieties, and L-functions, Vol. I (Ann Arbor, MI, 1988), Perspect. Math. Academic Press, Boston, MA.

KOTTWITZ, R. E. 1992. Points on some Shimura varieties over finite fields. *J. Amer. Math. Soc.* 5:373–444.

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.

LAN, K.-W. 2017. An example-based introduction to Shimura varieties. Preprint.

LANGLANDS, R. P. 1976. Some contemporary problems with origins in the Jugendtraum, pp. 401–418. *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.

LANGLANDS, R. P. AND RAPOPORT, M. 1987. Shimuravarietäten und Gerben. *J. Reine Angew. Math.* 378:113–220.

LEE, J. M. 1997. Riemannian manifolds, volume 176 of *Graduate Texts in Mathematics*. 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. 1979a. Etude d'une class d'isogénie, pp. 73–81. *In* Variétés de Shimura et fonctions L, volume 6 of *Publications Mathématiques de l'Université Paris VII*. Université de Paris VII U.E.R. de Mathématiques, Paris. Translation available at jmilne.org.

MILNE, J. S. 1979b. Points on Shimura varieties mod p, pp. 165–184. *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.

MILNE, J. S. 1983. The action of an automorphism of $\mathbf{C}$ on a Shimura variety and its special points, pp. 239–265. *In* Arithmetic and geometry, Vol. I. Birkhäuser Boston, Boston, MA.

MILNE, J. S. 1986. Abelian varieties, pp. 103–150. *In* Arithmetic geometry (Storrs, Conn., 1984). Springer, New York. A corrected version is available at jmilne.org.

MILNE, J. S. 1987. The (failure of the) Hasse principle for centres of semisimple groups. 6th June, 1987. Available at jmilne.org.

MILNE, J. S. 1990. Canonical models of (mixed) Shimura varieties and automorphic vector bundles, pp. 283–414. *In* Automorphic forms, Shimura varieties, and L-functions, Vol. I (Ann Arbor, MI, 1988), Perspect. Math. Academic Press, Boston, MA.

MILNE, J. S. 1992. The points on a Shimura variety modulo a prime of good reduction, pp. 151–253. *In* The zeta functions of Picard modular surfaces. Univ. Montréal, Montreal, QC.

MILNE, J. S. 1994a. Motives over finite fields, pp. 401–459. *In* Motives (Seattle, WA, 1991), volume 55 of *Proc. Sympos. Pure Math.* Amer. Math. Soc., Providence, RI.

MILNE, J. S. 1994b. Shimura varieties and motives, pp. 447–523. *In* Motives (Seattle, WA, 1991), Proc. Sympos. Pure Math. Amer. Math. Soc., Providence, RI.

MILNE, J. S. 1999. Descent for Shimura varieties. *Michigan Math. J.* 46:203–208.

MILNE, J. S. 2003. Gerbes and abelian motives. arXiv:math.AG/0301304.

MILNE, J. S. 2013. Shimura varieties and moduli, pp. 467–548. *In* Handbook of moduli. Vol. II, volume 25 of *Adv. Lect. Math. (ALM)*. Int. Press, Somerville, MA. The numbering used here follows that in the manuscript at jmilne.org, which differs from the published version.

MILNE, J. S. 2017. Algebraic Groups, the theory of group schemes of finite type over a field, volume 170 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

MILNE, J. S. AND SHIH, K.-Y. 1979. Shimura varieties: conjugates and the action of complex multiplication. October 1979, 154 pages. Published as the three articles (Annals 1981, LNM 900 1982). Available at jmilne.org.

MILNE, J. S. AND SHIH, K.-Y. 1981a. The action of complex conjugation on a Shimura variety. *Ann. of Math. (2)* 113:569–599.

MILNE, J. S. AND SHIH, K.-Y. 1981b. Automorphism groups of Shimura varieties and reciprocity laws. *Amer. J. Math.* 103:911–935.

MOK, N. AND ZHONG, J. Q. 1989. Compactifying complete Kähler-Einstein manifolds of finite topological type and bounded curvature. *Ann. of Math. (2)* 129:427–470.

MUMFORD, D. 1965. Geometric invariant theory. Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Band 34. Springer-Verlag, Berlin.

MUMFORD, D. 1970. Abelian varieties. Tata Institute of Fundamental Research Studies in Mathematics, No. 5. Published for the Tata Institute of Fundamental Research, Bombay, by Oxford University Press.

MUMFORD, D. 1977. Hirzebruch's proportionality theorem in the noncompact case. *Invent. Math.* 42:239–272.

NADEL, A. AND TSUJI, H. 1988. Compactification of complete Kähler manifolds of negative Ricci curvature. *J. Differential Geom.* 28:503–512.

ONO, T. 1959. On some arithmetic properties of linear algebraic groups. *Ann. of Math. (2)* 70:266–290.

PETERS, C. 2017a. On rigidity of locally symmetric spaces. arXiv:1704.03426.

PETERS, C. 2017b. Rigidity of spreadings and fields of definition. arXiv:1704.03410.

PFAU, M. 1996. The conjecture of Langlands and Rapoport for certain Shimura varieties of non-rational weight. *J. Reine Angew. Math.* 471:165–199.

PLATONOV, V. AND RAPINCHUK, A. 1994. Algebraic groups and number theory, volume 139 of *Pure and Applied Mathematics*. Academic Press Inc., Boston, MA.

PRASAD, G. AND RAPINCHUK, A. S. 2008. Developments on the congruence subgroup problem after the work of Bass, Milnor and Serre. arXiv:0809.1622. Published in the 5th volume of the collected works of Milnor.

RAGHUNATHAN, M. S. 1972. Discrete subgroups of Lie groups. Springer-Verlag, New York.

RAGHUNATHAN, M. S. 2004. The congruence subgroup problem. *Proc. Indian Acad. Sci. Math. Sci.* 114:299–308.

REIMANN, H. 1997. The semi-simple zeta function of quaternionic Shimura varieties, volume 1657 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin.

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, volume 4 of *Kanô Memorial Lectures*. Iwanami Shoten, Tokyo.

SATAKE, I. 2001. Compactifications, old and new [translation of Sūgaku **5**1 (1999), no. 2, 129–141. *Sugaku Expositions* 14:175–189.

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.

SERRE, J.-P. 1964. Sur les groupes de congruence des variétés abéliennes. *Izv. Akad. Nauk SSSR Ser. Mat.* 28:3–20.

SERRE, J.-P. 1966. *Algèbres de Lie semi-simples complexes*. W. A. Benjamin, Inc., New York and Amsterdam. Translated as *Complex semisimple Lie algebras*, Springer-Verlag, Berlin, 1987.

SERRE, J.-P. 1967a. Complex multiplication, pp. 292–296. *In* Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965). Thompson, Washington, D.C.

SERRE, J.-P. 1967b. Groupes de congruence (d'après H. Bass, H. Matsumoto, J. Mennicke, J. Milnor, C. Moore. *Séminaire Bourbaki, 1966/67, no. 330* . Œuvres II, 460–469.

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.

SHIH, K.-Y. 1978. Existence of certain canonical models. *Duke Math. J.* 45:63–66.

SHIMURA, G. 1963. On analytic families of polarized abelian varieties and automorphic functions. *Ann. of Math. (2)* 78:149–192.

SHIMURA, G. 1967. Construction of class fields and zeta functions of algebraic curves. *Ann. of Math. (2)* 85:58–159.

SHIMURA, G. 1971. On arithmetic automorphic functions, pp. 343–348. *In* Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 2. Gauthier-Villars, Paris.

SHIMURA, G. 1998. Abelian varieties with complex multiplication and modular functions, volume 46 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ.

SHIMURA, G. AND TANIYAMA, Y. 1961. Complex multiplication of abelian varieties and its applications to number theory, volume 6 of *Publications of the Mathematical Society of Japan*. The Mathematical Society of Japan, Tokyo. Shimura 1998 is an expanded version of this work, and retains most of the same numbering.

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, volume 9 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA.

STOTHERS, W. W. 1984. Level and index in the modular group. *Proc. Roy. Soc. Edinburgh Sect. A* 99:115–126.

SURY, B. 2003. The congruence subgroup problem, volume 24 of *Texts and Readings in Mathematics*. Hindustan Book Agency, New Delhi.

TAELMAN, L. 2017. Complex multiplication and Shimura stacks. arXiv:1707.01236.

TATE, J. T. 1967. Global class field theory, pp. 162–203. *In* Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965). Thompson, Washington, D.C.

TATE, J. T. 1968. Classes d'isogénie des variétés abéliennes sur un corps fini (d'après T. Honda). Séminaire Bourbaki: Vol. 1968/69, Expose 352.

TAYLOR, J. L. 2002. Several complex variables with connections to algebraic geometry and Lie groups, volume 46 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.

TITS, J. 1979. Reductive groups over local fields, pp. 29–69. *In* Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1, Proc. Sympos. Pure Math., XXXIII. Amer. Math. Soc., Providence, R.I.

VASIU, A. 1999. Integral canonical models of Shimura varieties of preabelian type. *Asian J. Math.* 3:401–518.

VOISIN, C. 2002. Théorie de Hodge et géométrie algébrique complexe, volume 10 of *Cours Spécialisés*. Société Mathématique de France, Paris. Translated as *Hodge theory and complex algebraic geometry* Cambridge University Press, 2002.

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. 1956a. The field of definition of a variety. *Amer. J. Math.* 78:509–524.

WEIL, A. 1956b. On the theory of complex multiplication. *In* Proceedings of the international symposium on algebraic number theory, Tokyo & Nikko, 1955, pp. 9–22, Tokyo. Science Council of Japan.

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, volume 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 MORRIS, D. 2015. Introduction to Arithmetic Groups. Deductive Press, CA. arXiv:math/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.

Index