AN EXAMPLE-BASED INTRODUCTION TO SHIMURA VARIETIES

KAI-WEN LAN

Abstract. In this introductory article, we will explain what Shimura varieties are and why they are useful, with an emphasis on examples. We will also describe various compactifications of Shimura varieties, again with an emphasis on examples, and summarize what we know about the integral models of Shimura varieties and their compactifications.

Contents

1. Introduction 3
2. Double coset spaces 4
2.1. Algebraic groups 4
2.2. Manifolds 6
2.3. Shimura data 9
2.4. Shimura varieties and their canonical models 10
3. Hermitian symmetric domains 13
3.1. The case of $\text{Sp}_{2n}(\mathbb{R})$ 13
3.1.1. Siegel upper half-spaces 13
3.1.2. Transitivity of action 14
3.1.3. Bounded realization 15
3.1.4. Interpretation as conjugacy classes 16
3.1.5. Moduli of polarized abelian varieties 16
3.2. The case of $U_{a,b}$ 18
3.2.1. Bounded realization 18
3.2.2. Unbounded realization 19
3.2.3. The special case where $a = b$: Hermitian upper half-spaces 20
3.2.4. Generalized Cayley transformations and transitivity of actions 20
3.2.5. Interpretation as conjugacy classes 21
3.3. The case of $\text{SO}^*_n$ 22
3.3.1. Unbounded realization, and interpretation as a conjugacy class 22
3.3.2. The special case where $n = 2k$: quaternion upper half-spaces 23
3.4. The case of $\text{SO}_{a,2}(\mathbb{R})$ 23
3.4.1. Projective coordinates 23

2010 Mathematics Subject Classification. Primary 14G35; Secondary 11G18.

Key words and phrases. locally symmetric varieties, Shimura varieties, examples.

The author was partially supported by the National Science Foundation under agreement No. DMS-1352216, and by an Alfred P. Sloan Research Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the author and do not necessarily reflect the views of these organizations.

This article will be published in the proceedings of the ETHZ Summer School on Motives and Complex Multiplication.
3.4.2. The case where $a \geq 1$ and $b = 2$
3.4.3. Connected components and spin groups
3.4.4. Interpretation as conjugacy classes
3.5. The case of $E_7$
3.5.1. Cayley numbers
3.5.2. An exceptional Jordan algebra over $\mathbb{R}$
3.5.3. A real Lie group with Lie algebra $\mathfrak{e}_7(-25)$
3.5.4. An octonion upper half-space
3.5.5. Bounded realization
3.6. The case of $E_6$
3.6.1. Bounded and unbounded realizations
3.6.2. A real Lie group of Lie algebra $\mathfrak{e}_6(-14)$
3.6.3. Interpretation as conjugacy classes
3.7. A brief summary
4. Rational boundary components and compactifications
4.1. Examples of rational boundary components
4.1.1. Cases with $D = \mathcal{H}$
4.1.2. Cases with $D = \mathcal{H} \times \mathcal{H}$
4.1.3. Cases with $D = \mathcal{H}_a$
4.1.4. Cases with $D = \mathcal{H}_{a,b}$, with $a \geq b$
4.1.5. Cases with $D = \mathcal{H}_{SO}^+$
4.1.6. Cases with $D^+ = \mathcal{H}_{SO}_{a,2}^+$, with $a \geq 2$
4.1.7. Cases with $D = \mathcal{H}_{E_6}$ or $D = \mathcal{H}_{E_7}$
4.2. Compactifications and algebraicity
4.2.1. Overview
4.2.2. $(\Gamma \backslash D)^{\text{min}} = \Gamma \backslash D^*$
4.2.3. $(\Gamma \backslash D)_{\text{tor}}$
4.2.4. $(\Gamma \backslash D)_{\text{bs}}$
4.2.5. $(\Gamma \backslash D)_{\text{RBS}}$
4.2.6. Models over number fields or their rings of integers
4.2.7. Algebraic-geometric definition of modular forms
4.2.8. Mixed Shimura varieties
5. Integral models
5.1. PEL-type cases
5.1.1. PEL datum
5.1.2. Smooth PEL moduli problems
5.1.3. PEL-type Shimura varieties
5.1.4. Toroidal and minimal compactifications of PEL moduli
5.1.5. Integral models not defined by smooth moduli
5.1.6. Langlands–Rapoport conjecture
5.2. Hodge-type and abelian-type cases
5.2.1. Hodge-type Shimura varieties
5.2.2. Abelian-type Shimura varieties
5.2.3. Integral models of these Shimura varieties
5.2.4. Integral models of toroidal and minimal compactifications
5.3. Beyond abelian-type cases
5.4. Beyond Shimura varieties
Acknowledgements
1. Introduction

Shimura varieties are generalizations of modular curves, which have played an important role in many recent developments of number theory. Just to mention a few examples with which this author is more familiar, Shimura varieties were crucially used in the proof of the local Langlands conjecture for $\text{GL}_n$ by Harris and Taylor (see [HT01]); in the proof of the Iwasawa main conjecture for $\text{GL}_2$ by Skinner and Urban (see [SU14]); in the construction of $p$-adic $L$-functions for unitary groups by Eischen, Harris, Li, and Skinner (see [HLS06] and [EHLS20]); in the construction of Galois representations (in the context of the global Langlands correspondence) for all cohomological automorphic representations of $\text{GL}_n$ by Harris, Taylor, Thorne, and this author (see [HLTT16]); and in the analogous but deeper construction for torsion cohomological classes of $\text{GL}_n$ by Scholze (see [Sch15]).

One of the reasons that Shimura varieties are so useful is because they carry two important kinds of symmetries—the Hecke symmetry and the Galois symmetry. Roughly speaking, the Hecke symmetry is useful for studying automorphic representations, the Galois symmetry is useful for studying Galois representations, and the compatibility between the two kinds of symmetries is useful for studying the relations between the two kinds of representations, especially in the context of the Langlands program.

However, the theory of Shimura varieties does not have a reputation of being easy to learn. There are important foundational topics such as the notion of Shimura data and the existence of canonical models, and courses or seminars on Shimura varieties often spend a substantial amount of the time on such topics. But learning these topics might shed little light on how Shimura varieties are actually used—in some of the important applications mentioned above, the “Shimura varieties” there were not even defined using any Shimura data or canonical models. In reality, it is still a daunting task to introduce people to this rich and multifaceted subject.

In this introductory article, we will experiment with a somewhat different approach. We still urge the readers to read the excellent texts such as [Del71b], [Del79], and [Mil05]. But we will not try to repeat most of the standard materials, such as variation of Hodge structures, the theory of canonical models, etc, which have already been explained quite well by many authors. We simply assume that people will learn them from some other sources. Instead, we will try to explain the scope of the theory by presenting many examples, and many special facts which are only true for these examples. Our experience is that such materials are rather scattered in the literature, but can be very helpful for digesting the abstract concepts, and for learning how to apply the general results in specific circumstances. We hope that they will benefit the readers in a similar way.

In this introductory article, we will experiment with a somewhat different approach. We still urge the readers to read the excellent texts such as [Del71b], [Del79], and [Mil05]. But we will not try to repeat most of the standard materials, such as variation of Hodge structures, the theory of canonical models, etc, which have already been explained quite well by many authors. We simply assume that people will learn them from some other sources. Instead, we will try to explain the scope of the theory by presenting many examples, and many special facts which are only true for these examples. Our experience is that such materials are rather scattered in the literature, but can be very helpful for digesting the abstract concepts, and for learning how to apply the general results in specific circumstances. We hope that they will benefit the readers in a similar way.

We will keep the technical level as low as possible, and present mainly simple-minded arguments that should be understandable by readers who are willing to see matrices larger than $2 \times 2$ ones. This is partly because this article is based on our notes for several introductory lectures targeting audiences of very varied backgrounds. The readers will still see a large number of technical phrases, often
without detailed explanations, but we believe that such phrases can be kept as keywords in mind and learned later. We will provide references when necessary, but such references do not have to be consulted immediately—doing that will likely disrupt the flow of reading and obscure the main points. (It is probably unrealistic and unhelpful to try to consult all of the references, especially the historical ones.) We hope that the readers will find these consistent with our goal of explaining the scope of the theory. There will inevitably be some undesirable omissions even with our limited goal, but we hope that our overall coverage will still be helpful.

Here is an outline of the article. In Section 2 we define certain double coset spaces associated with pairs of algebraic groups and manifolds, and introduce the notion of Shimura data. We also summarize the main results concerning the quasi-projectivity of Shimura varieties and the existence of their canonical models. In Section 3 we give many examples of Hermitian symmetric domains, which provide the connected components of the manifolds needed in the definition of Shimura data. Our examples actually exhaust all possibilities of them. In Section 4 we give many examples of rational boundary components and describe the corresponding minimal compactifications. We also summarize the properties of several kinds of useful compactifications. In Section 5 we explain how the so-called integral models of Shimura varieties are constructed. This is very sensitive to the types of Shimura varieties we consider, and we also take this opportunity to summarize many special facts and many recent developments for each type of them. For the convenience of the readers, we have also included an index at the end of the article.

2. Double coset spaces

2.1. Algebraic groups. Let \( \widehat{\mathbb{Z}} := \varprojlim (\mathbb{Z}/N\mathbb{Z}) \), \( \mathbb{A}^\infty := \widehat{\mathbb{Z}} \otimes \mathbb{Q} \), and \( \mathbb{A} := \mathbb{R} \times \mathbb{A}^\infty \). Suppose \( G \) is a reductive algebraic group over \( \mathbb{Q} \). Then there is a natural way to define the topological groups \( G(\mathbb{R}) \), \( G(\mathbb{A}^\infty) \), and \( G(\mathbb{A}) = G(\mathbb{R}) \times G(\mathbb{A}^\infty) \), with the topology given either as a restricted product (following Weil) (see [PR94, Sec. 5.1]), or as an affine scheme (of finite presentation) over the topological ring \( \mathbb{A} \) (following Grothendieck), and the two approaches coincide in this case (see [Con12] for a detailed discussion on this matter). We will not try to define and summarize all the needed properties of reductive linear algebraic groups (from textbooks such as [Bor91] and [Spr98]). Instead, we will just provide many examples of such groups.

Example 2.1.1 (general and special linear groups). For each ring \( R \) and each integer \( n \geq 0 \), we define \( \text{GL}_n(R) \) to be the group of invertible \( n \times n \) matrices with entries in \( R \), and define \( \text{SL}_n(R) \) to be the subgroup of \( \text{GL}_n(R) \) formed by matrices of determinant one. When \( n = 1 \), we have \( \text{GL}_1(R) = \mathbb{G}_m(R) := R^\times \). The assignments of \( \text{GL}_n(R) \) and \( \text{SL}_n(R) \) are functorial in \( R \), and in fact they are the \( R \)-points of the affine group schemes \( \text{GL}_n \) and \( \text{SL}_n \) over \( \mathbb{Z} \), respectively. The pullbacks of these group schemes to \( \mathbb{Q} \) are affine algebraic varieties serving as prototypical examples of reductive algebraic groups. Moreover, the pullback of \( \text{SL}_n \) to \( \mathbb{Q} \) is a prototypical example of a semisimple algebraic group.

For simplicity of exposition, we will often introduce a linear algebraic group \( G \) by describing it as the pullback to \( \mathbb{Q} \) of some explicitly given closed subgroup scheme of \( \text{GL}_m \), for some \( m \), over \( \mathbb{Z} \). By abuse of notation, we will often denote this group scheme over \( \mathbb{Z} \) by the same symbol \( G \), which is then equipped with a fixed choice of a faithful matrix representation \( \text{GL}_m \). The readers should not be too worried
about these technical terminologies concerning group schemes over \( \mathbb{Z} \). Concretely, they just mean we will explicitly represent the elements of \( G(R) \) as invertible \( m \times m \) matrices, for each ring \( R \), whose entries satisfy certain defining conditions given by algebraic equations that are compatible with all base ring extensions \( R \to R' \).

**Example 2.1.2 (symplectic groups).** Let \( n \geq 0 \) be any integer, and let \( 0_n \) and \( 1_n \) denote the zero and identity matrices of size \( n \). Consider the skew-symmetric matrix

\[
J_n := \begin{pmatrix} 0_n & 1_n \\ -1_n & 0_n \end{pmatrix}
\]

of size \( 2n \). Then the assignments

\[
\text{Sp}_{2n}(R) := \{ g \in \text{GL}_{2n}(R) : t^g J_n g = J_n \}
\]

and

\[
\text{GSp}_{2n}(R) := \{ (g, r) \in \text{GL}_{2n}(R) \times R^\times : t^g J_n g = r J_n \}
\]

are functorial in \( R \), and define closed subgroup schemes \( \text{Sp}_{2n} \) and \( \text{GSp}_{2n} \) of \( \text{GL}_{2n} \) and \( \text{GL}_{2n+1} \), respectively. Here we view \( \text{GL}_{2n} \times \mathbb{G}_m \) as a closed subgroup scheme of \( \text{GL}_{2n+1} \) by block-diagonally embedding the matrix entries. To better understand the meaning of these definitions, consider the alternating pairing

\[
\langle \cdot, \cdot \rangle : R^{2n} \times R^{2n} \to R
\]

defined by setting

\[
\langle x, y \rangle = t^x J_n y,
\]

for all \( x, y \in R^{2n} \) (written as vertical vectors). For each \( g \in \text{GL}_{2n}(R) \), we have

\[
\langle gx, gy \rangle = t^x t^g J_n gy.
\]

Therefore, \( g \) satisfies the condition \( t^g J_n g = J_n \) exactly when

\[
\langle gx, gy \rangle = \langle x, y \rangle,
\]

for all \( x, y \in R^{2n} \). That is, \( g \) preserves the above pairing \( \langle \cdot, \cdot \rangle \). Note that we have not used exactly what \( J_n \) is, except for its skew-symmetry. Similarly, \( (g, r) \in \text{GL}_{2n}(R) \times R^\times \) satisfies the condition \( t^g J_n g = r J_n \) exactly when

\[
\langle gx, gy \rangle = r \langle x, y \rangle,
\]

for all \( x, y \in R^{2n} \). That is, \( g \) preserves the above pairing \( \langle \cdot, \cdot \rangle \) up to the scalar factor \( r \). The assignment of \( r \) to \( (g, r) \) defines a homomorphism

\[
\nu : \text{GSp}_{2n}(R) \to \mathbb{G}_m(R),
\]

called the similitude character of \( \text{GSp}_{2n}(R) \), whose kernel is exactly \( \text{Sp}_{2n}(R) \), for each ring \( R \). We can define similar groups and homomorphisms by replacing \( J_n \) with any other matrices (with entries over \( \mathbb{Z} \), if we still want to define a group scheme over \( \mathbb{Z} \)). In fact, many people would prefer to define \( \text{Sp}_{2n} \) and \( \text{GSp}_{2n} \) using anti-diagonal matrices (just not a “block anti-diagonal matrix” like the above \( J_n = \begin{pmatrix} 0_n & 1_n \\ -1_n & 0_n \end{pmatrix} \)).
Example 2.1.12 (orthogonal groups and special orthogonal groups). Consider any integers $n, a, b \geq 0$, and consider the symmetric matrices $1_n$ and

$$(2.1.13) \quad 1_{a,b} := \begin{pmatrix} 1_a & \cdot \\ \cdot & -1_b \end{pmatrix},$$

with empty slots being filled up by zero matrices of suitable sizes. Then the assignments

$$(2.1.14) \quad O_n(R) := \{ g \in \text{GL}_n(R) : tgg = 1_n \},$$

$$(2.1.15) \quad SO_n(R) := O_n(R) \cap \text{SL}_n(R),$$

$$(2.1.16) \quad O_{a,b}(R) := \{ g \in \text{GL}_n(R) : t^g1_{a,b}g = 1_{a,b} \},$$

$$(2.1.17) \quad SO_{a,b}(R) := O_{a,b}(R) \cap \text{SL}_{a+b}(R)$$

are functorial in $R$, and define closed subgroup schemes $O_n$, $SO_n$, $O_{a,b}$, and $SO_{a,b}$ of $\text{GL}_n$, $\text{GL}_n$, $\text{GL}_{a+b}$, and $\text{GL}_{a+b}$, respectively. As in the case of $\text{Sp}_{2n}(R)$ above, $O_n(R)$ and $O_{a+b}(R)$ are the group of matrices preserving the symmetric bilinear pairings defined by $1_n$ and $1_{a,b}$, which have signatures $(n, 0)$ and $(a, b)$, respectively.

2.2. **Manifolds.** Let us fix a choice of $G$, and consider any manifold $D$ with a smooth transitive action of $G(\mathbb{R})$. For technical reasons, we shall consider only algebraic groups $G$ over $\mathbb{Q}$ each of whose connected components in the Zariski topology contains at least one rational point. This requirement is satisfied by every connected algebraic group, because the identity element is always rational, and also by all the examples in Section 2.1. Then $G(\mathbb{Q})$ is dense in $G(\mathbb{R})$ in the real analytic topology, by applying a special case of the weak approximation theorem as in [PR94, Sec. 7.3, Thm. 7.7] to the identity component of $G$ (namely, the connected component containing the identity element) in the Zariski topology, and by translations between this identity component and other connected components.

Example 2.2.1. For each integer $n \geq 0$, consider the manifold

$$(2.2.2) \quad \mathcal{S}_n := \{ X \in \text{Sym}_n(\mathbb{R}) : X > 0, \det(X) = 1 \}$$

where $M_n$ (resp. $\text{Sym}_n$) denotes the space of $n \times n$ matrices (resp. symmetric matrices), and where $X > 0$ means being positive definite. Consider the action of $g \in \text{SL}_n(\mathbb{R})$ on $\mathcal{S}_n$ defined by

$$(2.2.3) \quad X \mapsto gX^t g,$$

for each $X \in \mathcal{S}_n$. This action is transitive because each positive definite matrix $X$ is of the form

$$(2.2.4) \quad X = Y^2 = Y^t Y$$

for some positive definite matrix $Y$, and if $\det(X) = 1$ then $\det(Y) = 1$ as well. The stabilizer of $X = 1_n \in \mathcal{S}_n$ is, by definition, $SO_n(\mathbb{R})$, and so we have

$$(2.2.5) \quad \mathcal{S}_n = \text{SL}_n(\mathbb{R}) \cdot 1_n \cong \text{SL}_n(\mathbb{R})/SO_n(\mathbb{R}).$$

Note that

$$(2.2.6) \quad \dim_{\mathbb{R}}(\mathcal{S}_n) = \frac{1}{2}n(n+1) - 1.$$

In the classification in [Hel01, Ch. X, Sec. 6], $\mathcal{S}_n$ is a noncompact Riemannian symmetric space of type $\Lambda I$. 
Example 2.2.7. When $n = 2$, we also have the familiar example of
\begin{equation}
\begin{aligned}
g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{R})
\end{aligned}
\end{equation}
acting on the Poincaré upper half-plane
\begin{equation}
\mathcal{H} := \{ z \in \mathbb{C} : \text{Im}(z) > 0 \}
\end{equation}
by the Möbius transformation
\begin{equation}
\begin{aligned}
z \mapsto gz = \frac{az + b}{cz + d},
\end{aligned}
\end{equation}
for each $z \in \mathcal{H}$. Note that this is actually induced by the natural \( \text{SL}_2(\mathbb{C}) \) action on \( \mathbb{P}^1(\mathbb{C}) \): If we identify $z \in \mathbb{C}$ with \( \begin{pmatrix} z \\ 1 \end{pmatrix} \in \mathbb{P}^1(\mathbb{C}) \), then
\begin{equation}
\begin{aligned}
\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z \\ 1 \end{pmatrix} = \begin{pmatrix} az + b \\ cz + d \end{pmatrix} \sim \begin{pmatrix} g z \\ 1 \end{pmatrix}.
\end{aligned}
\end{equation}
For readers who have studied modular forms, the factor \((cz + d)\) involved in the above identification between projective coordinates is exactly the same automorphy factor \((cz + d)\) in the definition of holomorphic modular forms. (This is not just a coincidence.) It is well known that this action of \( \text{SL}_2(\mathbb{R}) \) on \( \mathcal{H} \) is transitive (which will be generalized in Section 3.1.2 below), and the stabilizer of $i \in \mathcal{H}$ is
\begin{equation}
\begin{aligned}
\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \in \text{SL}_2(\mathbb{R}) \} = \text{SO}_2(\mathbb{R}) = \left\{ \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} : \theta \in \mathbb{R} \right\},
\end{aligned}
\end{equation}
so that we have
\begin{equation}
\begin{aligned}
\mathcal{H} = \text{SL}_2(\mathbb{R}) \cdot i \cong \text{SL}_2(\mathbb{R})/\text{SO}_2(\mathbb{R}).
\end{aligned}
\end{equation}

Example 2.2.14. The isomorphisms \( \mathcal{H} \cong \text{SL}_2(\mathbb{R})/\text{SO}_2(\mathbb{R}) \cong S_2 \) show that \( \mathcal{H} \) can be viewed as a special case of \( S_n \) with \( n = 2 \). But this is the only example of \( S_n \), with \( n \geq 1 \), such that \( S_n \) has a complex structure. Since \( \dim_\mathbb{C}(S_n) = \frac{1}{2} n(n + 1) - 1 \) is odd when \( n = 3 \), for example, there is no hope for \( S_3 \) to be a complex manifold.

For any \((G, D)\) as above, and for any open compact subgroup \( \mathcal{U} \) of \( G(\mathbb{A}^\infty) \), we can define the double coset space
\begin{equation}
\begin{aligned}
X_\mathcal{U} := G(\mathbb{Q}) \backslash (D \times G(\mathbb{A}^\infty))/\mathcal{U},
\end{aligned}
\end{equation}
where \( G(\mathbb{Q}) \) acts diagonally on \( D \times G(\mathbb{A}^\infty) \) from the left-hand side, and where \( \mathcal{U} \) only acts on \( G(\mathbb{A}^\infty) \) from the right-hand side. The reader might naturally wonder why we need to consider a complicated double quotient as in (2.2.15). One important justification is that the group \( G(\mathbb{A}^\infty) \) has a natural right action on the collection \( \{ X_\mathcal{U} \}_\mathcal{U} \), induced by
\begin{equation}
\begin{aligned}
D \times G(\mathbb{A}^\infty) \xrightarrow{\sim} D \times G(\mathbb{A}^\infty) : (x, h) \mapsto (x, hg),
\end{aligned}
\end{equation}
for each \( g \in G(\mathbb{A}^\infty) \), which maps \( X_{\mathcal{U} g \mathcal{U}^{-1}} \) to \( X_\mathcal{U} \) because \( h(gu g^{-1})g = hgu \), for all \( h \in G(\mathbb{A}^\infty) \) and \( u \in \mathcal{U} \). Such an action provides natural Hecke actions on, for example, the limit of cohomology groups \( \lim_{\mathcal{U} \to} H^*(X_\mathcal{U}, \mathbb{C}) \). Such a symmetry of \( \{ X_\mathcal{U} \}_\mathcal{U} \) is what we meant by Hecke symmetry in the introduction (see Section 1). This is crucial for relating the geometry of such double coset spaces to the theory of automorphic representations.
Let $D^+$ be a connected component of $D$, which admits a transitive action of $G(\mathbb{R})^+$, the identity component (namely, the connected component containing the identity element) of $G(\mathbb{R})$ in the real analytic topology. Let $G(\mathbb{R})_+$ denote the stabilizer of $D^+$ in $G(\mathbb{R})$, which necessarily contains $G(\mathbb{R})^+$. Let

$$G(\mathbb{Q})_+ := G(\mathbb{Q}) \cap G(\mathbb{R})_+, \quad (2.2.17)$$

which is the subgroup of $G(\mathbb{Q})$ stabilizing $D^+$. Then

$$G(\mathbb{Q})_+ \backslash (D^+ \times G(\mathbb{A}^\infty))/\mathcal{U} \to G(\mathbb{Q}) \backslash (D \times G(\mathbb{A}^\infty))/\mathcal{U}, \quad (2.2.18)$$

is surjective because $G(\mathbb{Q})$ is dense in $G(\mathbb{R})$ (see the beginning of this Section 2.2), and is injective by definition. It is known (see [Bor63, Thm. 5.1]) that

$$\#(G(\mathbb{Q})_+ \backslash G(\mathbb{A}^\infty)/\mathcal{U}) < \infty, \quad (2.2.19)$$

which means there exists a subset $\{g_i\}_{i \in I}$ of $G(\mathbb{A}^\infty)$ indexed by a finite set $I$ such that we have a disjoint union

$$G(\mathbb{A}^\infty) = \coprod_{i \in I} G(\mathbb{Q})_+ g_i \mathcal{U}. \quad (2.2.20)$$

Then

$$X_{\mathcal{U}} \cong G(\mathbb{Q})_+ \backslash (D^+ \times G(\mathbb{A}^\infty))/\mathcal{U}$$

$$= \coprod_{i \in I} G(\mathbb{Q})_+ \backslash (D^+ \times G(\mathbb{Q})_+ g_i \mathcal{U})/\mathcal{U}$$

$$= \coprod_{i \in I} \Gamma_i \backslash D^+, \quad (2.2.21)$$

where

$$\Gamma_i := G(\mathbb{Q})_+ \cap (g_i \mathcal{U} g_i^{-1}), \quad (2.2.22)$$

because $\gamma g_i u = g_i u$ exactly when $\gamma \in g_i \mathcal{U} g_i^{-1}$, for any $\gamma \in G(\mathbb{Q})_+$, $i \in I$, and $u \in \mathcal{U}$.

Each such $\Gamma_i$ is an arithmetic subgroup of $G(\mathbb{Q})$; namely, a subgroup commensurable with $G(\mathbb{Z})$ in the sense that $\Gamma_i \cap G(\mathbb{Z})$ has finite indices in both $\Gamma_i$ and $G(\mathbb{Z})$. In fact, since $\mathcal{U}$ is an open compact subgroup of $G(\mathbb{A}^\infty)$, each such $\Gamma_i$ is a congruence subgroup of $G(\mathbb{Z})$; namely, a subgroup containing the principal congruence subgroup $\ker(G(\mathbb{Z}) \to G(\mathbb{Z}/N\mathbb{Z}))$ for some integer $N \geq 1$. By definition, congruence subgroups are arithmetic subgroups. Note that, although the definition of principal congruence subgroups depends on the choice of some faithful matrix representation $G \hookrightarrow \text{GL}_m$ over $\mathbb{Z}$ (see Section 2.1), the definition of congruence subgroups does not, because congruence subgroups of $G(\mathbb{Q})$ can be characterized alternatively as the intersections of $G(\mathbb{Q})$ with open compact subgroups of $G(\mathbb{A}^\infty)$. More generally, the definition of arithmetic subgroups of $G(\mathbb{Q})$ does not depend on the choice of the faithful matrix representation either (see [Bor69, 7.13] or [PR94, Sec. 4.1]).

If each $\Gamma_i$ acts freely on $D^+$, then $X_{\mathcal{U}}$ is a manifold because $D^+$ is. This is the case when each $\Gamma_i$ is a neat arithmetic subgroup of $G(\mathbb{Q})$ (as in [Bor69, (17.1)]), which is in turn the case when $\mathcal{U}$ is a neat open compact subgroup of $G$ (as in [Pin89, 0.6]). For reading the remainder of this article, it suffices to know that, when people say such groups are neat (or, much less precisely, sufficiently small), they just want to ensure that the stabilizers of geometric actions are all trivial. For simplicity, we shall tacitly assume that the arithmetic subgroups or open compact subgroups we will encounter are all neat, unless otherwise stated.
Now the question is what each $\Gamma_i \backslash D^+$ is, after knowing that it is a manifold because $\Gamma_i$ is neat. What additional structures can we expect from such quotients?

**Example 2.2.23.** When $G = \text{SL}_2$ and $D = \mathcal{H}$ as in Example 2.2.7, the groups $\Gamma_i$ are congruence subgroups of $\text{SL}_2(\mathbb{Z})$, each of which contains the principal congruence subgroup

$$\Gamma(N) := \ker(\text{SL}_2(\mathbb{Z}) \to \text{SL}_2(\mathbb{Z}/N\mathbb{Z}))$$

for some integer $N \geq 1$. The quotients $\Gamma_i \backslash \mathcal{H}$ are familiar objects called *modular curves*, each of which admits a good compactification into a *compact Riemann surface* (or, in other words, *complete algebraic curves* over $\mathbb{C}$) by adding a finite number of points, which are called *cusps*. (We will revisit this example in more detail in Section 4.1.1 below.)

**Remark 2.2.25.** We may replace the manifold $D$ above with other geometric objects over $D$ that still have a smooth (but not necessarily transitive) action of $G(R)$. For example, when $G = \text{SL}_2$ and $D = \mathcal{H}$, we may consider the space $\mathcal{H} \times \text{Sym}^k(\mathbb{C}^2)$, and obtain natural vector bundles or local systems over $X_U$.

### 2.3. Shimura data.

Shimura varieties are not just any double coset spaces $X_U$ as above. For arithmetic applications, it is desirable that each $X_U$ is an algebraic variety over $\mathbb{C}$ with a model over some canonically determined number field.

This led to the notion of a *Shimura datum* $(G, D)$ (see [Del79, 2.1.1] and [Mil05, Def. 5.5]). Consider the Deligne torus $S$, which is the restriction of scalars $\text{Res}_{\mathbb{C}/\mathbb{R}} G_{m, \mathbb{C}}$ (as an algebraic group over $\mathbb{R}$), so that $S(\mathbb{R}) = \mathbb{C}^\times$. Then we require $G$ to be a *connected reductive algebraic group*, and require $D$ to be a $G(\mathbb{R})$-conjugacy class of homomorphisms

$$h : S \to G_{\mathbb{R}},$$

where the subscript “$\mathbb{R}$” means “base change to $\mathbb{R}$” as usual, satisfying the following conditions:

1. The representation defined by $h$ and the *adjoint representation* of $G(\mathbb{R})$ on $\mathfrak{g} = \text{Lie} G(\mathbb{C})$ induces a decomposition

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}^+ \oplus \mathfrak{p}^-$$

such that

$$z \in S(\mathbb{R}) = \mathbb{C}^\times$$

acts by $1$, $z/\bar{z}$, and $\bar{z}/z$, respectively, on the three summands.

2. $h(i)$ induces a *Cartan involution* on $G_{\text{ad}}(\mathbb{R})$. (Here $G_{\text{ad}}$ denotes the adjoint quotient of $G$; namely, the quotient of $G$ by its center.)

3. $G_{\mathbb{R}}^\text{ad}$ has no nontrivial $\mathbb{Q}$-simple factor $H$ such that $H(\mathbb{R})$ is compact (or, equivalently, such that the composition of $h$ with the projection to $H$ is trivial, because of the previous condition (2)).

In this case, any finite dimensional $\mathbb{R}$-representation of $G(\mathbb{R})$ defines a *variation of (real) Hodge structures*, and $D$ is a finite union of *Hermitian symmetric domains*. Without going into details, let us just emphasize that $D$ and its quotients $X_U$, for neat open compact subgroups $U$ of $G(\mathbb{A}^\infty)$, are then *complex manifolds*, not just real ones.

For studying only the connected components $D^+$ (of some $D$ which appeared in the last paragraph) and their quotients, it is easier to work with what is called a
connected Shimura datum \((G, D^+)\) (see [Mil05, Def. 4.22]), which requires \(G\) to be a connected semisimple algebraic group, and requires \(D^+\) to be a \(G^{\text{ad}}(\mathbb{R})^+\)-conjugacy class of homomorphisms

\[
  h : \mathbb{S} \to G^{\text{ad}}_{\mathbb{R}},
\]

where \(G^{\text{ad}}(\mathbb{R})^+\) denotes (as before) the identity component of \(G^{\text{ad}}(\mathbb{R})\) in the real analytic topology, satisfying the same three conditions (1), (2), and (3) as above.

A homomorphism as in (2.3.4) determines and is determined by a homomorphism

\[
  h : U_1 \to G^{\text{ad}}_{\mathbb{R}},
\]

where

\[
  U_1 := \{ z \in \mathbb{C} : |z| = 1 \}
\]

is viewed as an algebraic group over \(\mathbb{R}\) here, via the canonical homomorphism

\[
  \mathbb{S} \to U_1
\]

induced by \(z \mapsto z/\bar{z}\). In many cases we will study, the restriction of the homomorphism (2.3.4) to \(U_1\) (now viewed as a subgroup of \(\mathbb{S}\)) lifts to a homomorphism

\[
  h : U_1 \to G_{\mathbb{R}},
\]

whose composition with \(G_{\mathbb{R}} \to G^{\text{ad}}_{\mathbb{R}}\) then provides a square-root of the \(h\) in (2.3.5).

This is nonstandard, but sometimes more convenient, when we prefer to describe \(G_{\mathbb{R}}\) instead of \(G^{\text{ad}}_{\mathbb{R}}\) in many practical examples (to be given in Section 3 below).

2.4. Shimura varieties and their canonical models. Now we are ready to state the following beautiful results, whose assertions can be understood without knowing anything about their proofs:

**Theorem 2.4.1** (Satake, Baily–Borel, and Borel; see [BB66] and [Bor72]). Suppose \((G, D)\) is a Shimura datum as in Section 2.3. Then the whole collection of complex manifolds \(\{X_U\}_U\), with \(U\) varying among neat open compact subgroups of \(G(A^{\infty})\), is the complex analytification (see [Ser56, §2]) of a canonical collection of smooth quasi-projective varieties over \(C\). Moreover, the analytic covering maps

\[
  X_U \to X_{U'},
\]

when \(U \subset U'\), are also given by the complex analytifications of canonical finite étale algebraic morphisms between the corresponding varieties.

Alternatively, suppose \((G, D^+)\) is a connected Shimura datum as in Section 2.3. Then analogous assertions hold for the collection \(\{\Gamma \backslash D^+ \}_\Gamma\), with \(\Gamma\) varying among neat arithmetic subgroups of \(G^{\text{ad}}(\mathbb{Q})\) that are contained in \(G^{\text{ad}}(\mathbb{R})^+\). (If we consider also \((G, D)\), where \(D\) is the corresponding \(G(\mathbb{R})\)-conjugacy class of any \(h\) in \(D^+\), then the connected components of \(X_U\), where \(U\) is a neat open compact subgroup of \(G(A^{\infty})\), are all of the form \(\Gamma \backslash D^+\), as explained in Section 2.2.)

**Theorem 2.4.3** (Shimura, Deligne, Milne, Borovoi, and others; see [Shi70], [Del71b], [Del79], [Mil83], [Bor84], and [Bor87]). Suppose \((G, D)\) is a Shimura datum as in Section 2.3. Then there exists a number field \(F_0\), given as a subfield of \(\mathbb{C}\) depending only on \((G, D)\), called the reflex field of \((G, D)\), such that the whole collection of complex manifolds \(\{X_U\}_U\), with \(U\) varying among neat open compact subgroups of \(G(A^{\infty})\), is the complex analytification of the pullback to \(\mathbb{C}\) of a canonical collection of smooth quasi-projective varieties over \(F_0\), which satisfies
certain additional properties qualifying them as the **canonical models** of \( \{X_U\}_U \). Moreover, the analytic covering maps

\[
X_U \to X_{U'},
\]

when \( U \subset U' \), are also given by the complex analytifications of canonical finite étale algebraic morphisms defined over \( F_0 \) between the corresponding canonical models.

**Remark 2.4.5.** Theorem \[2.4.1\] is proved by constructing the so-called Satake–Baily–Borel or minimal compactifications of \( X_U \) or \( \Gamma \backslash D^+ \), which are projective varieties over \( \mathbb{C} \). (See Sections \[4.2.2\] and \[4.2.7\] below.) The assertion that the quasi-projective varieties \( X_U \) and \( \Gamma \backslash D^+ \) are defined over \( \mathbb{Q} \), the algebraic closure of \( \mathbb{Q} \) in \( \mathbb{C} \), can be shown without using the theory of canonical models, as in \[Fal84\]. But these general results cannot explain why each \( X_U \) is defined over some particular number field.

**Remark 2.4.6.** In Theorem \[2.4.3\] since each \( X_U \) is algebraic and defined over \( F_0 \), its arithmetic invariants such as the étale cohomology group \( H^0(X_U, \mathbb{Q}_l) \), for any prime number \( l \), carries a canonical action of \( \text{Gal}(\overline{Q}/F_0) \), where the algebraic closure \( \overline{Q} \) of \( Q \) in \( C \) contains \( F_0 \) because \( F_0 \) is given as a subfield of \( C \). (Recall that any base change between separably closed fields induces a canonical isomorphism between the étale cohomology groups—see, for example, \[Del77\], Arcata, V, 3.3.) This is an instance of what we meant by Galois symmetry in the introduction (see Section \[1\]). Since the maps \( X_U \to X_{U'} \) are also algebraic and defined over \( F_0 \), we have a canonical action of \( \text{Gal}([\overline{Q}/F_0]) \) on \( \lim_{U \to U'} \prod H^0(X_U, \mathbb{Q}_l) \), which is compatible with the Hecke action of \( G(\mathbb{A}^\infty) \). This is an instance of what we meant by the compatibility between Hecke and Galois symmetries in the introduction (see Section \[1\]).

**Remark 2.4.7.** In Theorem \[2.4.3\] if we assume instead that \( (G, D^+) \) is a connected Shimura datum, and consider the collection \( \{\Gamma \backslash D^+\}_\Gamma \) as in Theorem \[2.4.1\] then it is only true that each \( \Gamma \backslash D^+ \) is defined over a number field depending on both \( (G, D^+) \) and \( \Gamma \), but in general we cannot expect the whole collection \( \{\Gamma \backslash D^+\}_\Gamma \), or just the subcollection formed by those \( \Gamma \) whose preimage in \( G(\mathbb{Q}) \) are congruence subgroups, to be defined over a single number field depending only on \( (G, D) \). This is a generalization of the following classical phenomenon: When \( (G, D) = (\text{SL}_2, \mathcal{H}) \), and when interpreting the quotient \( \Gamma(N)\backslash \mathcal{H} \) as in Example \[2.2.23\] as a parameter space for complex elliptic curves \( E_z = \mathbb{C}/(\mathbb{Z}z + \mathbb{Z}) \) with level \( N \)-structures

\[
\alpha_N : (\mathbb{Z}/N\mathbb{Z})^2 \to E_z[N] : (a, b) \mapsto (\frac{a}{\mathcal{H}z} + \frac{b}{\mathcal{H}}) \mod (\mathbb{Z}z + \mathbb{Z}) ,
\]

where \( z \in \mathcal{H} \) and \( E_z[N] \) denotes the \( N \)-torsion subgroup of \( E_z \), by identifying \( E_z \) with \( E_{\gamma z} \) for all \( \gamma \in \Gamma(N) \), the value of the Weil pairing of \( \alpha_N((1, 0), (0, 1)) \) is a constant primitive \( N \)-th root \( \zeta_N \) of unity on \( \Gamma(N)\backslash \mathcal{H} \). (See the introduction of \[DR73\].) We can only expect \( \Gamma(N)\backslash \mathcal{H} \) to have a model over \( \mathbb{Q}(\zeta_N) \), rather than \( \mathbb{Q} \).

**Remark 2.4.9.** In the context of Theorem \[2.4.3\] people may refer to either the varieties \( X_U \) over \( C \), or their canonical models over \( F_0 \), or the projective limits of such varieties over either \( C \) or \( F_0 \), as the **Shimura varieties** associated with the Shimura datum \( (G, D) \). For the sake of clarity, people sometimes say that \( X_U \) and its canonical model are Shimura varieties at level \( U \). In the context of Remark \[2.4.7\] people often refer to quotients of the form \( \Gamma\backslash D^+ \), where \( \Gamma \) are arithmetic
subgroups of
\[(2.4.10) \quad G^\text{ad}(\mathbb{Q})^+ := G^\text{ad}(\mathbb{Q}) \cap G^\text{ad}(\mathbb{R})^+\]
whose preimages under the canonical homomorphism \(G(\mathbb{Q}) \to G^\text{ad}(\mathbb{Q})\) are congruence subgroups of \(G(\mathbb{Q})\), as the connected Shimura varieties associated with the connected Shimura datum \((G, D^+)\). We allow all such subgroups \(\Gamma\) of \(G^\text{ad}(\mathbb{Q})^+\), rather than only the images of the congruence subgroups of \(G(\mathbb{Q})\), so that the connected Shimura varieties thus defined are useful for studying all connected components of Shimura varieties.

Remark 2.4.11. By a morphism of Shimura data
\[(2.4.12) \quad (G_1, D_1) \to (G_2, D_2),\]
we mean a group homomorphism \(G_1 \to G_2\) mapping \(D_1\) into \(D_2\). If \(D_1 = G_1(\mathbb{R}) \cdot h_0\) is the conjugacy class of some homomorphism \(h_0 : \mathbb{S} \to G_{1,\mathbb{R}}\), then this means the composition of \(h_0\) with \(G_1(\mathbb{R}) : \mathbb{S} \to G_{1,\mathbb{R}}\), if \(D_2\) lies in the conjugacy class \(D_2\). If \(U_1\) and \(U_2\) are open compact subgroups of \(G_1(\mathbb{A}^\infty)\) and \(G_2(\mathbb{A}^\infty)\), respectively, such that \(U_1\) is mapped into \(U_2\), then we obtain the corresponding morphism \(X_{U_1} \to X_{U_2}\) between Shimura varieties. This morphism is defined over the subfield of \(\mathbb{C}\) generated by the reflex fields of \((G_1, D_1)\) and \((G_2, D_2)\). There are also analogues of these assertions for connected Shimura data and the corresponding connected Shimura varieties.

Remark 2.4.13. An important special case of a morphism of Shimura data as in Remark 2.4.11 is when the homomorphism \(G_1 \to G_2\) is injective, and when \(U_1\) is exactly the pullback of \(U_2\) under the induced homomorphism \(G_1(\mathbb{A}^\infty) \to G_2(\mathbb{A}^\infty)\).
In this case, up to replacing \(U_2\) with an open compact subgroup (satisfying the same conditions), \(X_{U_1}\) is a closed subvariety of \(X_{U_2}\). (It is always possible to find such a \(U_2\) for any given \(U_1\) — see [Del71b, Prop. 1.15].) The neatness of \(U_2\) can be arranged when \(U_1\) is neat, by taking \(U_2\) to be generated by the image of \(U_1\) and some suitable neat open compact subgroup of \(G_2(\mathbb{A}^\infty)\).) Subvarieties of this kind are called special subvarieties. The most important examples are given by special points or (with some additional assumptions on the Shimura data) CM points, which are zero-dimensional special subvarieties defined by subgroups \(G_1\) of \(G_2\) that are tori (see [Mil05, Def. 12.5]). These are generalizations of the points of modular curves parameterizing CM elliptic curves with level structures, which are represented by points of the Poincaré upper half-plane with coordinates in imaginary quadratic field extensions of \(\mathbb{Q}\). Zero-dimensional Shimura varieties and special points are important because their canonical models can be defined more directly, and hence they are useful for characterizing the canonical models of Shimura varieties of positive dimensions.

We will not try to explain the proofs of Theorems 2.4.1 and 2.4.3 in this introductory article. In fact, we will not even try to explain in any depth the background notions such as variations of Hodge structures or Hermitian symmetric domains, or key notions such as zero-dimensional Shimura varieties, the Shimura reciprocity law, and the precise meaning of canonical models, because most existing texts already present them quite well, and there is little we can add to them. Instead, we will start by presenting many examples of \((G, D)\) and \((G, D^+)\), so that the readers can have some idea to where the theory applies. We hope this will benefit not only readers who would like to better understand the existing texts, but also readers who have concrete arithmetic or geometric problems in mind.
3. HERMITIAN SYMMETRIC DOMAINS

In this section, we will explore many examples of Hermitian symmetric domains, which can be interpreted as the generalizations of the Poincaré upper half-plane $\mathcal{H}$ (see Example 2.2.7) in many different ways. Most of these will give the $(G(\mathbb{R}), \mathcal{D})$ or $(G(\mathbb{R}), \mathcal{D}^+)$ for some Shimura data $(G, \mathcal{D})$ or connected Shimura data $(G, \mathcal{D}^+)$, respectively, as in Section 2.3, although we will not be using the structure of $G(\mathbb{Q})$ in this section. (Later we will indeed use the structure of $G(\mathbb{Q})$ in Section 4.)

However, we will not define Hermitian symmetric domains. Our excuse for this omission is that there is a complete classification of all irreducible ones (namely, those that are not isomorphic to nontrivial products of smaller ones), and we will see all of them in our examples. (For a more formal treatment, there are general introductions such as [Mil05, Sec. 1], with references to more advanced texts such as [Hel01].)

In our personal experience, especially in number-theoretic applications requiring actual calculations, the abstract theory of Hermitian symmetric domains cannot easily replace the knowledge of these explicit examples. (Already in the classical theory for $\text{SL}_2$ or $\text{GL}_2$, it is hard to imagine studying modular forms without ever introducing the Poincaré upper half-plane.)

We shall begin with the case of $\text{Sp}_{2n}(\mathbb{R})$ in Section 3.1, followed by the case of $\text{U}_{a,b}$ in Section 3.2, the case of $\text{SO}^*_{2n}$ in Section 3.3, the case of $\text{SO}_{a,2}(\mathbb{R})$ in Section 3.4, the case of $\text{E}_7$ in Section 3.5, the case of $\text{E}_6$ in Section 3.6, and a summary of all these cases in Section 3.7.

3.1. The case of $\text{Sp}_{2n}(\mathbb{R})$.

3.1.1. Siegel upper half-spaces. Let $n \geq 0$ be any integer. Consider the Siegel upper half-space

$$\mathcal{H}_n := \{ Z \in \text{Sym}_n(\mathbb{C}) : \text{Im}(Z) > 0 \} = \{ Z \in \mathbb{M}_n(\mathbb{C}) : {^{t\text{f}}} Z - Z = 0, \frac{1}{2} \left( {^t\text{f}} Z - Z \right) < 0 \} = \left\{ Z \in \mathbb{M}_n(\mathbb{C}) : {^t}\left( \begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array} \right) \left( \begin{array}{c} Z \\ 1 \end{array} \right) = 0, \frac{1}{2} {^t}\left( \begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array} \right) \left( \begin{array}{c} Z \\ 1 \end{array} \right) < 0 \right\},$$

where $\mathbb{M}_n$ (resp. $\text{Sym}_n$) denotes the space of $n \times n$ matrices (resp. symmetric matrices) as before, and where the notation $> 0$ (resp. $< 0$) means positive definiteness (resp. negative definiteness) of matrices, which admits a natural (left) action of $g = \left( \begin{array}{cc} A & B \\ C & D \end{array} \right) \in \text{Sp}_{2n}(\mathbb{R})$:

$$Z \mapsto gZ := (AZ + B)(CZ + D)^{-1}$$

(note that $(CZ + D)^{-1}$ is multiplied to the right of $(AZ + B)$). Again, such an action should be understood in generalized projective coordinates:

$$(A B) \left( \begin{array}{c} Z \\ 1 \end{array} \right) = \left( \begin{array}{cc} AZ + B \\ CZ + D \end{array} \right) \sim \left( \begin{array}{c} gZ \\ 1 \end{array} \right)$$

(by multiplying $(CZ + D)^{-1}$ to the right).

Remark 3.1.1.4. By generalized projective coordinates given by an $m \times n$ matrix, we mean an $m \times n$ matrix of rank $n$ that is identified up to the right action of invertible
$n \times n$ matrices. Giving such a matrix is equivalent to defining a point of the Grassmannian of $n$-dimensional subspaces of an $m$-dimensional vector space. Later we will encounter some similar generalized projective coordinates when defining other Hermitian symmetric domains, and the readers can interpret them as defining points on some other Grassmannians. This is closely related to the interpretation of Hermitian symmetric domains as parameter spaces of variations of Hodge structures, where the Grassmannians parameterize some Hodge filtrations.

Let us briefly explain why the action is indeed defined. Since

$$t\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

we have

$$t\begin{pmatrix} Z \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} Z \\ 1 \end{pmatrix} = t\begin{pmatrix} Z \\ 1 \end{pmatrix} g\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} g\begin{pmatrix} Z \\ 1 \end{pmatrix}$$

if and only if

$$t\begin{pmatrix} gZ \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} gZ \\ 1 \end{pmatrix} = 0;$$

and

$$\frac{1}{\pi} t\begin{pmatrix} Z \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} Z \\ 1 \end{pmatrix} = \frac{1}{\pi} t\begin{pmatrix} Z \\ 1 \end{pmatrix} g\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} g\begin{pmatrix} Z \\ 1 \end{pmatrix}$$

if and only if

$$\frac{1}{\pi} t\begin{pmatrix} gZ \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} gZ \\ 1 \end{pmatrix} < 0$$

if and only if

$$\frac{1}{\pi} t\begin{pmatrix} gZ \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} gZ \\ 1 \end{pmatrix} < 0.$$

Thus, $Z \in \mathcal{H}_n$ if and only if $gZ \in \mathcal{H}_n$, and the action is indeed defined.

3.1.2. Transitivity of action. If $Z = X + iY \in \mathcal{H}_n$, then

$$\begin{pmatrix} 1 & -X \\ 0 & 1 \end{pmatrix} Z = iY,$$

with $Y > 0$, and so there exists some $A \in \text{GL}_n(\mathbb{R})$ such that

$$tAYA = 1_n,$$

which shows that

$$\begin{pmatrix} tA & 0 \\ 0 & A \end{pmatrix} (iY) = i1_n.$$

These show that

$$\mathcal{H}_n = \text{Sp}_{2n}(\mathbb{R}) \cdot (i1_n).$$

The stabilizer of $i1_n$ is

$$K := \left\{ \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \in \text{Sp}_{2n}(\mathbb{R}) \right\} \cong U_n := \left\{ g = A + iB : t\overline{g} = 1_n \right\}.$$
AN EXAMPLE-BASED INTRODUCTION TO SHIMURA VARIETIES 15

(where the conditions on \(A\) and \(B\) are the same for \(\begin{pmatrix} A & B \\ -B & A \end{pmatrix} \in \text{Sp}_{2n}(\mathbb{R})\) and for \(g = A + iB \in U_n\), which is a maximal compact subgroup of \(\text{Sp}_{2n}(\mathbb{R})\). (Certainly, the precise matrix expressions of these elements depend on the choice of the point \(i1_n\).) Thus, we have shown that

\[(3.1.2.6) \quad \mathcal{H}_n \cong \text{Sp}_{2n}(\mathbb{R})/U_n.\]

When \(n = 1\), this specializes to the isomorphism \(\mathcal{H} = \mathcal{H}_1 \cong \text{SL}_2(\mathbb{R})/\text{SO}_2(\mathbb{R})\) in Example 2.2.7 because \(\text{SL}_2(\mathbb{R}) \cong \text{Sp}_2(\mathbb{R})\) and \(\text{SO}_2(\mathbb{R}) \cong U_1\). Note that

\[(3.1.2.7) \quad \dim \mathcal{H}_n = \dim \text{Sym}_n(\mathbb{C}) = \frac{1}{2}n(n+1), \quad \text{for all } n \geq 0.\]

3.1.3. Bounded realization. There is also a bounded realization of \(\mathcal{H}_n\). Note that

\[(3.1.3.1) \quad \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} i \\ 1 \end{pmatrix} \sim \begin{pmatrix} 0 \\ 2i \end{pmatrix}.\]

Set

\[(3.1.3.2) \quad U := (\mathbb{Z} - i1_n)(\mathbb{Z} + i1_n)^{-1},\]

so that

\[(3.1.3.3) \quad \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} Z \\ 1 \end{pmatrix} = \begin{pmatrix} Z - i1_n \\ Z + i1_n \end{pmatrix} \sim \begin{pmatrix} U \\ 1 \end{pmatrix}\]

in generalized projective coordinates (cf. Remark 3.1.1.4). Note that this assignment maps \(Z = i1_0\) to \(U = 0\). Since

\[(3.1.3.4) \quad \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} = \begin{pmatrix} 0 & 2i \\ -2i & 0 \end{pmatrix} = 2i \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\]

we have

\[(3.1.3.5) \quad ^tZ = Z \iff ^tU = U,\]

and

\[(3.1.3.6) \quad \frac{1}{2i} ({}^tZ - Z) = \frac{1}{2i} \begin{pmatrix} Z \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} Z \\ 1 \end{pmatrix} = \begin{pmatrix} Z \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -i \end{pmatrix} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} Z \\ 1 \end{pmatrix} < 0\]

\[\iff {}^tUU - 1_n = \begin{pmatrix} U \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} U \\ 1 \end{pmatrix} < 0,\]

and so

\[(3.1.3.7) \quad \mathcal{H}_n \cong \mathcal{D}_n := \{ U \in \text{Sym}_n(\mathbb{C}) : ^tUU - 1_n < 0 \}.\]

This generalizes the Cayley transforms that move the Poincaré upper half-plane \(\mathcal{H}\) to the open unit disc \(\{ u \in \mathbb{C} : |u| < 1 \}\). (But the formula for the action of \(\text{Sp}_{2n}(\mathbb{R})\) has to be conjugated, and is more complicated.)

In the classification in [Hel01, Ch. X, Sec. 6], \(\mathcal{H}_n \cong \mathcal{D}_n\) is a Hermitian symmetric domain of type C I. In Cartan’s classification, it is a bounded symmetric domain of type III_\(n\).
3.1.4. Interpretation as conjugacy classes. Now consider the homomorphism

\[ h_0 : U_1 \to \operatorname{Sp}_{2n}(\mathbb{R}) : x + yi \mapsto \begin{pmatrix} x & -y \\ y & x \end{pmatrix}. \]

For simplicity, in the matrix at the right-hand side, we denoted by \( x \) and \( y \) the scalar multiples of the \( n \times n \) identity matrix \( 1_n \) by \( x \) and \( y \), respectively. (We will adopt similar abuses of notation later, without further explanation.) Then we have

\[ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \]

\[ \iff \begin{pmatrix} -B & A \\ -D & C \end{pmatrix} = \begin{pmatrix} C & D \\ -A & -B \end{pmatrix} \]

\[ \iff A = D, B = -C, \]

which shows that

\[ \text{Cent}_{h_0}(\operatorname{Sp}_{2n}(\mathbb{R})) = U_n \]

as subgroups of \( \operatorname{Sp}_{2n}(\mathbb{R}) \), where Cent denotes the centralizer, and so that

\[ \mathcal{H}_n = \operatorname{Sp}_{2n}(\mathbb{R}) \cdot h_0, \]

with the action given by conjugation. That is, \( \mathcal{H}_n \) is the \( \operatorname{Sp}_{2n}(\mathbb{R}) \)-conjugacy class of \( h_0 \). In this case, \( (\operatorname{Sp}_{2n}, \mathcal{H}_n) \) is a connected Shimura datum as in Section 2.3 (where we use \( \operatorname{Sp}_{2n} \) to denote the algebraic group over \( \mathbb{Q} \), not the group scheme over \( \mathbb{Z} \)), taking into account the variant (2.3.8) of (2.3.5) as explained there. (For simplicity, we shall not repeat such an explanation in what follows.)

The above is compatible with the similarly defined action of \( \operatorname{GSp}_{2n}(\mathbb{R}) \) on

\[ \mathcal{H}_n^\pm := \{ Z \in M_n(\mathbb{C}) : \text{Im}(Z) > 0 \text{ or } \text{Im}(Z) < 0 \} \]

Then

\[ \mathcal{H}_n^\pm = \operatorname{GSp}_{2n}(\mathbb{R}) \cdot h_0, \]

where

\[ h_0 : \mathbb{C}^\times \to \operatorname{GSp}_{2n}(\mathbb{R}) : x + yi \mapsto \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \]

also encodes the action of \( \mathbb{R}^\times \subset \mathbb{C}^\times \), and is better for the purpose of studying variations of Hodge structures (and also for various other reasons). In this case, \( (\operatorname{GSp}_{2n}, \mathcal{H}_n^\pm) \) is a Shimura datum as in Section 2.3 (where we use \( \operatorname{GSp}_{2n} \) to denote the algebraic group over \( \mathbb{Q} \), not the group scheme over \( \mathbb{Z} \)).

3.1.5. Moduli of polarized abelian varieties. For each \( Z \in \mathcal{H}_n \), we have a lattice

\[ L_Z := \mathbb{Z}^n Z + \mathbb{Z}^n \]

in \( \mathbb{C}^n \) such that

\[ A_Z := \mathbb{C}^n / L_Z \]

is not only a complex torus, but also a polarized abelian variety. (See [Mum70] Sec. 1–3 and [Mum83] Ch. II, Sec. 1 and 4.) When \( n = 1 \), this is just an elliptic curve.

The point is that the abelian variety \( A_Z = \mathbb{C}^n / L_Z \) (as a complex torus) can be identified with the fixed real torus

\[ \mathbb{R}^{2n} / \mathbb{Z}^{2n} \]
with its complex structure induced by the one on $\mathbb{R}^{2n}$ defined by
\[
(3.1.5.4) \quad h(i) \in \text{Sp}_{2n}(\mathbb{R}) \subset \text{GL}_{2n}(\mathbb{R}),
\]
where the homomorphism
\[
(3.1.5.5) \quad h = g \cdot h_0 \in \mathcal{H}_n = \text{Sp}_{2n}(\mathbb{R}) \cdot h_0
\]
corresponds to the point
\[
(3.1.5.6) \quad Z = g \cdot (i1_n) \in \mathcal{H}_n = \text{Sp}_{2n}(\mathbb{R}) \cdot (i1_n).
\]
Accordingly, varying the lattice $L_Z$ in $\mathbb{C}^n$ with $Z = g \cdot (i1_n)$ corresponds to varying the complex structure $J = h(i)$ on $\mathbb{R}^{2n}$ with $h = g \cdot h_0$. (Such a variation of complex structures can be interpreted as a variation of Hodge structures.)

The pairing on $\mathbb{Z}^{2n}$ defined by the perfect pairing $\langle \cdot, \cdot \rangle$ defines (up to a sign convention) a principal polarization $\lambda_Z$ on $A_Z$ (and the existence of such a polarization was the reason that the complex torus $A_Z$ is an abelian variety). If $(A_Z, \lambda_Z) \cong (A_{Z'}, \lambda_{Z'})$ for $Z$ and $Z'$ corresponding to $h$ and $h'$, respectively, then we have an induced isomorphism $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ preserving $\mathbb{Z}^{2n}$ and $\langle \cdot, \cdot \rangle$, which defines an element $g = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \text{Sp}_{2n}(\mathbb{Z})$ conjugating $h$ to $h'$ and mapping $Z$ to $Z'$. Conversely, if we have
\[
(3.1.5.7) \quad Z' = gZ = (AZ + B)(CZ + D)^{-1}
\]
for some $g = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \text{Sp}_{2n}(\mathbb{Z})$, then we have a homothety of lattices
\[
(3.1.5.8) \quad L_Z = \mathbb{Z}^{2n}Z + \mathbb{Z}^{2n} = \mathbb{Z}^{2n}(AZ + B) + \mathbb{Z}^{2n}(CZ + D)
\]
\[
= (\mathbb{Z}^{2n}Z' + \mathbb{Z}^{2n})(CZ + D) = L_{Z'}(CZ + D) \sim L_{Z'},
\]
which implies that $(A_Z, \lambda_Z) \cong (A_{Z'}, \lambda_{Z'})$. If $g$ lies in a congruence subgroup $\Gamma$ of $\text{Sp}_{2n}(\mathbb{Z})$, then the above isomorphism $(A_Z, \lambda_Z) \cong (A_{Z'}, \lambda_{Z'})$ also respect certain (symplectic) level structures. Therefore, $\Gamma \backslash \mathcal{H}_n$ (or the manifolds $X_U$ defined using $G = \text{GSp}_{2n}$ and $D = \mathcal{H}_n^\pm$) are Siegel modular varieties parameterizing polarized abelian varieties with certain level structures (depending on $\Gamma$ or $U$). The spaces $\Gamma \backslash \mathcal{H}_n$ are, a priori, just complex manifolds. The name “Siegel modular varieties” will be justified in Sections 4.1.3 and 4.2.2 below, where we shall see that they are indeed the complex analytifications of some canonical algebraic varieties.

The fact that spaces like $\Gamma \backslash \mathcal{H}_n$ parameterize polarized abelian varieties with level structures is perhaps just a coincidence, but such a coincidence is extremely important. One can redefine them as the $\mathbb{C}$-points of certain moduli problems of polarized abelian schemes with level structures; and then their constructions can be extended over the integers in number fields, because the moduli schemes of polarized abelian schemes, called Siegel moduli schemes, have been constructed over the integers. (See [MFK94] Ch. 7 for the construction based on geometric invariant theory. Alternatively, geometric invariant theory can be avoided by first constructing algebraic spaces or Deligne–Mumford stacks by verifying Artin’s criterion—see [Art69] and [Lan13] Sec. B.3 for such a criterion, and see [FC90] Ch. I, Sec. 4] and [Lan13] Ch. 2 for its verification for the moduli of polarized abelian schemes; and then showing that they are quasi-projective schemes by constructing their minimal compactifications as in [FC90] Ch. V] or [Lan13] Ch. 7]—see Section 5.1.4 below.)
Such integral models of $\Gamma \backslash \mathcal{H}_n$ are very useful because we can understand their reductions in positive characteristics by studying polarized abelian varieties over finite fields, for which we have many powerful tools. We will encounter more general cases of $\Gamma \backslash \mathcal{D}$ later, but we have good methods for constructing integral models for them (namely, methods for which we can say something useful about the outputs) only in cases that can be related to Siegel modular varieties at all.

3.2. The case of $U_{a,b}$. We learned some materials here from [Shi98] and [Shi78].

3.2.1. Bounded realization. Let $a \geq b \geq 0$ be any integers. Recall that we have introduced the symmetric matrix $1_{a,b} = \begin{pmatrix} 1_a & -1_b \\ 1_b & 1_b \end{pmatrix}$ in (2.1.13), which we now view as a Hermitian matrix, which defines a Hermitian pairing of signature $(a, b)$. Consider the group

$$U_{a,b} := \{ g \in \text{GL}_{a+b}(\mathbb{C}) : \begin{pmatrix} 1_a & -1_b \\ 1_b & 1_b \end{pmatrix} g = 1_{a,b} \},$$

which acts on

$$D_{a,b} := \{ U \in M_{a,b}(\mathbb{C}) : \begin{pmatrix} U & 1_a \\ 1_b & 1_b \end{pmatrix} \begin{pmatrix} 1_a & -1_b \\ 1_b & 1_b \end{pmatrix} = t^tUU - 1_b < 0 \},$$

where $M_{a,b}$ denotes the space of $a \times b$ matrices, by

$$U \mapsto gU = (AU + B)(CU + D)^{-1}$$

for each $g = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in U_{a,b}$. This action is based on the identity

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} U & 1_a \\ 1_b & 1_b \end{pmatrix} = \begin{pmatrix} AU + B \\ CU + D \end{pmatrix} \sim \begin{pmatrix} (AU + B)(CU + D)^{-1} & 1_a \\ 1_b & 1_b \end{pmatrix}\begin{pmatrix} 1_a & -1_b \\ 1_b & 1_b \end{pmatrix}$$

in generalized projective coordinates (cf. Remark 3.1.1.4). Note that $U_{a,b}$ is a real Lie group, but not a complex one, despite the use of complex numbers in its coordinates; and that

$$\dim_{\mathbb{C}} D_{a,b} = \dim_{\mathbb{C}} (M_{a,b}(\mathbb{C})) = ab.$$

For any $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \in U_{a,b}$, as soon as $B = 0$, we must also have $C = 0$, by the defining condition in (3.2.1.1). Hence, the stabilizer of $0 \in D_{a,b}$ is

$$K := \left\{ \begin{pmatrix} A \\ D \end{pmatrix} \in U_{a,b} \right\} \cong U_a \times U_b : \begin{pmatrix} A \\ D \end{pmatrix} \mapsto (A, D).$$

Remark 3.2.1.7. For the purpose of introducing Hermitian symmetric domains, it might be more natural to consider the semisimple Lie group $SU_{a,b}$—namely, the subgroup of $U_{a,b}$ of elements of determinant one—or its adjoint quotient, but we still consider the reductive Lie group $U_{a,b}$, because it is easier to write down homomorphisms of the form $U_1 \to U_{a,b}$, as in (2.3.8).

However, to show the transitivity of the action of $U_{a,b}$ on $D_{a,b}$, and for many other computations for which it is easier to work with block upper-triangular matrices, it will be easier to switch to some unbounded realization.
3.2.2. Unbounded realization. Consider
\[(3.2.2.1) \quad U'_{a,b} := \{ g \in \text{GL}_{a+b}(\mathbb{C}) : \ 'gJ_{a,b}g = J_{a,b} \}, \]
where the matrix
\[(3.2.2.2) \quad J_{a,b} := \begin{pmatrix} 1_b & S \\ -1_b & 0 \end{pmatrix} \]
is skew-Hermitian with \(-iS > 0\), for some choice of a skew-Hermitian matrix \(S\) (depending on the context), so that there exists some \(T \in \text{GL}_n(\mathbb{C})\) such that \(-iS = 'TT\). We warn the readers that the notation of \(J_{a,b}\) and \(U'_{a,b}\) here is rather nonstandard. Then \(U'_{a,b}\) acts on
\[(3.2.2.3) \quad H_{a,b} := \left\{ \begin{pmatrix} Z \\ W \end{pmatrix} \in M_{a,b}(\mathbb{C}) \cong M_b(\mathbb{C}) \times M_{a-b,b}(\mathbb{C}) : \right. \]
\[\left. -i'J_{a,b} \begin{pmatrix} Z \\ W \end{pmatrix} = -i('Z - Z + 'WSW) < 0 \right\} \]
by
\[(3.2.2.4) \quad \begin{pmatrix} Z \\ W \end{pmatrix} \mapsto g \begin{pmatrix} Z \\ W \end{pmatrix} = \begin{pmatrix} (AZ + EW + B)(CZ + HW + D)^{-1} \\ (FZ + MW + G)(CZ + HW + D)^{-1} \end{pmatrix}, \]
for each \(g = \begin{pmatrix} A & E & B \\ F & M & G \\ C & H & D \end{pmatrix} \in U'_{a,b}\), based on the identity
\[(3.2.2.5) \quad g \begin{pmatrix} Z \\ W \end{pmatrix} = \begin{pmatrix} (AZ + EW + B)(CZ + HW + D)^{-1} \\ (FZ + MW + G)(CZ + HW + D)^{-1} \end{pmatrix}, \]
in generalized projective coordinates (cf. Remark 3.1.1.4).

If we consider
\[(3.2.2.6) \quad \text{Herm}_b(\mathbb{C}) := \{ X \in M_b(\mathbb{C}) : 'X = X \}, \]
where \(\text{Herm}_b\) denotes the space of \(b \times b\) Hermitian matrices (which are over \(\mathbb{C}\) here), then we have a canonical isomorphism
\[(3.2.2.7) \quad M_b(\mathbb{C}) \cong \text{Herm}_b(\mathbb{C}) \otimes \mathbb{R} : X \mapsto \text{Re}(X) + i\text{Im}(X), \]
where
\[(3.2.2.8) \quad \text{Re}(X) := \frac{1}{2}(X + 'X) \]
and
\[(3.2.2.9) \quad \text{Im}(X) := \frac{1}{2i}(X - 'X) \]
are the Hermitian real and imaginary parts (abusively denoted by the usual symbols), and we can rewrite the condition
\[(3.2.2.10) \quad -i('Z - Z + 'WSW) < 0 \]
as
\[(3.2.2.11) \quad \text{Im}(Z) > \frac{1}{2} 'W(-iS)W. \]
3.2.3. The special case where \( a = b \): Hermitian upper half-spaces. In this case, we can omit the second coordinate \( W \) in the above unbounded realization, and we have

\[
\mathcal{H}_{b,b} = \{ Z \in M_b(\mathbb{C}) \cong \text{Herm}_b(\mathbb{C}) \otimes \mathbb{C} : \text{Im}(Z) > 0 \},
\]

which can be viewed as a generalization of

\[
\mathcal{H}_n = \{ Z \in \text{Sym}_n(\mathbb{C}) \cong \text{Sym}_n(\mathbb{R}) \otimes \mathbb{C} : \text{Im}(Z) > 0 \}.
\]

That is, we can view the symmetric matrices in \( \text{Sym}_n(\mathbb{R}) \) as being Hermitian with respect to the trivial involution of \( \mathbb{R} \), which then generalizes to the usual complex Hermitian matrices in \( \text{Herm}_n(\mathbb{C}) \) as being Hermitian with respect to the complex conjugation of \( \mathbb{C} \) over \( \mathbb{R} \). These two examples will be further generalized with \( \mathbb{R} \) and \( \mathbb{C} \) replaced with the other two normed (but possibly nonassociative) division algebras \( \mathbb{H} \) and \( \mathbb{O} \) over \( \mathbb{R} \), the rings of Hamilton and Cayley numbers, respectively, in Sections 3.3.2 and 3.5.4 below.

3.2.4. Generalized Cayley transformations and transitivity of actions. Since

\[
-\frac{i}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ \sqrt{2} & -\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 \ 1 \\ -1 \ 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \ T^{-1} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \ T^{-1} \frac{i}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 \ 1 \\ 1 \ -1 \end{pmatrix},
\]

we have an isomorphism

\[
U'_{a,b} \simeq U_{a,b} : g \mapsto \begin{pmatrix} \frac{1}{\sqrt{2}} \ T^{-1} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \ T^{-1} \frac{i}{\sqrt{2}} \end{pmatrix} g \begin{pmatrix} \frac{1}{\sqrt{2}} \ T^{-1} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \ T^{-1} \frac{i}{\sqrt{2}} \end{pmatrix},
\]

and accordingly an isomorphism

\[
\mathcal{H}_{a,b} \simeq \mathcal{D}_{a,b} : \begin{pmatrix} Z \\ W \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{\sqrt{2}} \ T^{-1} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \ T^{-1} \frac{i}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} Z \\ W \end{pmatrix} \sim \begin{pmatrix} U \\ 1 \end{pmatrix},
\]

equivariant with the above isomorphism \(\text{(3.2.4.2)}\). This is a generalization of the classical Cayley transformation and maps \( (i1_b,0) \in \mathcal{H}_{a,b} \) to \( 0 \in \mathcal{D}_{a,b} \).

Now, given any \( \begin{pmatrix} Z \\ W \end{pmatrix} \in \mathcal{H}_{a,b} \), we have

\[
\begin{pmatrix} 1 & * & * \\ 1 & -W & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} Z \\ W \\ 1 \end{pmatrix} = \begin{pmatrix} Z' \\ 0 \\ 1 \end{pmatrix},
\]

\[
\begin{pmatrix} 1 & -\text{Re}(Z') \\ 1 & 1 \end{pmatrix} \begin{pmatrix} Z' \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} i\text{Im}(Z') \\ 0 \\ 1 \end{pmatrix},
\]

\[
\begin{pmatrix} iA^{-1} \\ 1 \end{pmatrix} \begin{pmatrix} i\text{Im}(Z') \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} i1_b \\ 0 \\ 1 \end{pmatrix},
\]

\[
\begin{pmatrix} tA \\ 1 \end{pmatrix} \begin{pmatrix} i\text{Im}(Z') \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} i1_b \\ 0 \\ 1 \end{pmatrix},
\]

\[
\begin{pmatrix} tA \\ 1 \end{pmatrix} \begin{pmatrix} i\text{Im}(Z') \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} i1_b \\ 0 \\ 1 \end{pmatrix},
\]

\[
\begin{pmatrix} iA^{-1} \\ 1 \end{pmatrix} \begin{pmatrix} i\text{Im}(Z') \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} i1_b \\ 0 \\ 1 \end{pmatrix}.
\]
for some $Z' \in M_b(\mathbb{C})$ and some $A \in \text{GL}_b(\mathbb{C})$ such that $\text{Im}(Z') = \mathcal{M}A$, where the first matrix in each equation is some element of $U'_{a,b}$. These show that

$$H_{a,b} = U'_{a,b} \cdot \begin{pmatrix} i1_b \\ 0 \end{pmatrix} \cong \mathcal{D}_{a,b} = U_{a,b} \cdot 0 \cong U_{a,b} / (U_a \times U_b),$$

and hence the actions of $U_{a,b}$ and $U'_{a,b}$ on $\mathcal{D}_{a,b}$ and $H_{a,b}$, respectively, are transitive.

In the classification in [Hej01, Ch. X, Sec. 6], $H_{a,b} \cong \mathcal{D}_{a,b}$ is a Hermitian symmetric domain of type A III. In Cartan’s classification, it is a bounded symmetric domain of type $1_{ab}$.

### 3.2.5. Interpretation as conjugacy classes.

Consider the homomorphism

$$h_0 : U_1 \rightarrow U_{a,b} : x + yi \mapsto \begin{pmatrix} x - yi \\ x + yi \end{pmatrix},$$

with $x - yi$ and $x + yi$ denoting multiples of $1_a$ and $1_b$, respectively. Then

$$\text{Cent}_{h_0}(U_{a,b}) = U_a \times U_b$$

(as subgroups of $U_{a,b}$) and so that

$$\mathcal{D}_{a,b} = U_{a,b} \cdot h_0.$$

Via the inverse of the isomorphism $U_{a,b} \cong U'_{a,b}$ above defined by conjugation, we obtain

$$h_0' : U_1 \rightarrow U'_{a,b} : x + yi \mapsto \begin{pmatrix} x \\ y \\ x - yi \\ -y \end{pmatrix},$$

so that

$$H_{a,b} = U'_{a,b} \cdot h_0'.$$

Both of these are compatible with the extensions of the actions of $U_{a,b}$ and $U'_{a,b}$ on $\mathcal{D}_{a,b}$ and $H_{a,b}$, respectively, to some actions of

$$\begin{align*}
GU_{a,b} := & \{ (g, r) \in \text{GL}_{a+b}(\mathbb{C}) \times \mathbb{R}^\times : \mathcal{M}1_{a,b}g = r1_{a,b} \} \\
GU'_{a,b} := & \{ (g, r) \in \text{GL}_{a+b}(\mathbb{C}) \times \mathbb{R}^\times : \mathcal{M}J_{a,b}g = rJ_{a,b} \}
\end{align*}$$

on some

$$\begin{align*}
\mathcal{D}_{a,b}^\pm := & GU_{a,b} \cdot h_0 \\
\mathcal{H}_{a,b}^\pm := & GU_{a,b} \cdot h_0'
\end{align*}$$

(where we omit the explicit descriptions of $\mathcal{D}_{a,b}^\pm$ and $\mathcal{H}_{a,b}^\pm$, for simplicity), where $h_0$ and $h_0'$ are now homomorphisms

$$h_0 : \mathbb{C}^\times \rightarrow GU_{a,b}$$

and

$$h_0' : \mathbb{C}^\times \rightarrow GU'_{a,b}$$

defined by the same expressions as in (3.2.5.1) and (3.2.5.4), respectively. (As in Section 3.1.4, the upshot here is that the pairs $(GU_{a,b}, \mathcal{D}_{a,b}^\pm)$ and $(GU'_{a,b}, \mathcal{H}_{a,b}^\pm)$ can
be of the form \((G(\mathbb{R}), \mathcal{D})\) for some Shimura data \((G, \mathcal{D})\). We shall omit such remarks in later examples.)

Since the real part of any skew-Hermitian pairing is alternating, there is an injective homomorphism of the form

\[(3.2.5.12) \quad U_{a,b} \cong U'_{a,b} \hookrightarrow \text{Sp}_{2n}(\mathbb{R}),\]

with \(n = a + b\), whose pre-composition with \(h_0 : U_1 \rightarrow U_{a,b}\) lies in the conjugacy class of the analogous homomorphism for \(\text{Sp}_{2n}(\mathbb{R})\) in \([3.1.4.1]\).

3.3. The case of \(\text{SO}^*_2n\).

3.3.1. Unbounded realization, and interpretation as a conjugacy class. Consider the subgroup \(\text{SO}^*_2n\) of

\[(3.3.1.1) \quad \text{SO}_2n(\mathbb{C}) = \{g \in \text{SL}_2n(\mathbb{C}) : {}^tgg = 1_{2n}\}\]

preserving the skew-Hermitian form defined by \((-1_n 1_n)^t\). That is,

\[(3.3.1.2) \quad \text{SO}^*_2n = \text{SO}_2n(\mathbb{C}) \cap U'_{n,n}\]

as subgroups of \(\text{GL}_{2n}(\mathbb{C})\), where \(U'_{n,n}\) is defined as in \([3.2.2.1]\), with Lie algebra

\[(3.3.1.3) \quad \text{so}^*_2n = \left\{ \begin{pmatrix} A & B \\ -B & A \end{pmatrix} : A, B \in M_n(\mathbb{C}), {}^tA = -A, {}^tB = B \right\}\]

as Lie subalgebras of \(M_{2n}(\mathbb{C})\) (see, for example, [Kna01, Ch. I, Sec. 1] or [Kna02, Ch. I, Sec. 17]). By explanations as before, \(\text{SO}^*_2n\) acts transitively on

\[(3.3.1.4) \quad \mathcal{H}_{\text{SO}^*_2n} := \left\{ \begin{pmatrix} Z \in M_n(\mathbb{C}) : {}^t\begin{pmatrix} Z & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} Z \\ 1 \end{pmatrix} &= {}^tZZ + 1 = 0, \\ -i^t\begin{pmatrix} Z \\ 1 \end{pmatrix} \begin{pmatrix} Z & 1 \\ -1 & 1 \end{pmatrix} &= -({}^tZ - Z) < 0 \right\},\]

whose stabilizer at \(i1_n \in \mathcal{H}_{\text{SO}^*_2n}\) is

\[(3.3.1.5) \quad K := \left\{ \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \in \text{SO}^*_2n : A, B \in M_n(\mathbb{R}) \right\} \cong U_n,\]

so that

\[(3.3.1.6) \quad \mathcal{H}_{\text{SO}^*_2n} = \text{SO}^*_2n \cdot h_0 \cong \text{SO}^*_2n / U_n,\]

where

\[(3.3.1.7) \quad h_0 : U_1 \rightarrow \text{SO}^*_2n : x + yi \mapsto \begin{pmatrix} x & -y \\ y & x \end{pmatrix}.\]

As before, there is also another version of \(\text{SO}^*_2n\) which acts on some bounded realization \(\mathcal{D}_{\text{SO}^*_2n}\), but we shall omit them for simplicity. Since \(\text{SO}^*_2n \subset U'_{n,n}\), by the same explanation as in Section [3.2.5], there are injective homomorphisms

\[(3.3.1.8) \quad \text{SO}^*_2n \hookrightarrow \text{Sp}_{4n}(\mathbb{R}),\]

whose pre-composition with \(h_0 : U_1 \rightarrow \text{SO}^*_2n\) lies in the conjugacy class of the analogous homomorphism for \(\text{Sp}_{4n}(\mathbb{R})\) as in \([3.1.4.1]\).

In the classification in [Hel01, Ch. X, Sec. 6], \(\mathcal{H}_{\text{SO}^*_2n} \cong \mathcal{D}_{\text{SO}^*_2n}\) is a Hermitian symmetric domain of type \(D\) III. In Cartan’s classification, it is a bounded symmetric domain of type \(\Pi_1\).
3.3.2. The special case where \( n = 2k \): quaternion upper half-spaces. Since we have the isomorphism

\[
H \cong \mathbb{C}^2 : x + yi + j(z + wi) \mapsto \begin{pmatrix} x + yi \\ z + wi \end{pmatrix}
\]

of right \( \mathbb{C} \)-modules, where \( H \) denotes the Hamiltonian quaternion algebra over \( \mathbb{R} \), we have a homomorphism

\[
H \to M_2(\mathbb{C}) : x + yi + j(z + wi) \mapsto \begin{pmatrix} x + yi & -z + wi \\ z + wi & x - yi \end{pmatrix}
\]

induced by the left action of \( H \) on itself. Using this, we can show that

\[
SO^*_4k \cong \{ g \in GL_{2k}(H) : t_g \left( \begin{smallmatrix} 1_k \\ -1_k \end{smallmatrix} \right) g = \left( \begin{smallmatrix} 1_k \\ -1_k \end{smallmatrix} \right) \},
\]

namely, the subgroup \( Sp_{2k}(H) \) of \( GL_{2k}(H) \) respecting the skew-Hermitian pairing on \( H_k \) defined by \( \left( \begin{smallmatrix} 1_k \\ -1_k \end{smallmatrix} \right) \) (see, for example, [GW09, Sec. 1.1.4], combined with a Gram–Schmidt process as in [Lan13, Sec. 1.2.3]); and that

\[
H_{SO^*_4k} \cong \{ Z \in \text{Herm} \mathbb{C} \otimes \mathbb{R} : \text{Im}(Z) > 0 \},
\]

with the action of \( g = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in GL_{2k}(H) \) given by

\[
gZ = (AZ + B)(CZ + D)^{-1},
\]

which is similar to the action of \( Sp_{2k}(\mathbb{R}) \) on \( H_k \) in Section 3.1.1. This is a quaternion upper half-space, further generalizing the examples in Sections 3.1.1 and 3.2.3.

3.4. The case of \( SO_{a,b}(\mathbb{R}) \). We learned some materials here from [Sat80, Appendix, Sec. 6].

3.4.1. Projective coordinates. Let \( a, b \geq 0 \) be any integers. Consider

\[
SO_{a,b}(\mathbb{R}) := \{ g \in \text{SL}_{a+b}(\mathbb{R}) : t^g 1_{a,b} g = 1_{a,b} \},
\]

where \( 1_{a,b} = \begin{pmatrix} 1_a \\ -1_b \end{pmatrix} \) is defined as before. Let us temporarily write \( V := \mathbb{R}^{a+b} \) and \( V_{\mathbb{C}} := V \otimes \mathbb{C} \cong \mathbb{C}^{a+b} \). Then, by definition, \( SO_{a,b}(\mathbb{R}) \) acts on

\[
\tilde{H}_{SO_{a,b}} := \{ v \in V_{\mathbb{C}} : t^v \begin{pmatrix} 1_a \\ -1_b \end{pmatrix} v = 0, t^v \begin{pmatrix} 1_a \\ -1_b \end{pmatrix} v < 0 \} / \mathbb{C}^x.
\]

Note that \( v \neq 0 \) in the above definition, and hence the quotient by \( \mathbb{C}^x \) means we are actually working with a subset of \( \mathbb{P}^{a+b-1}(\mathbb{C}) \cong (V_{\mathbb{C}} - \{ 0 \}) / \mathbb{C}^x \). (This is consistent with our use of generalized projective coordinates in previous examples.) However, \( \tilde{H}_{SO_{a,b}} \) is in general not connected, and even its connected components do not have the structure of Hermitian symmetric domains for arbitrary \( a, b \geq 0 \).
3.4.2. The case where \(a \geq 1\) and \(b = 2\). In this case, we do get Hermitian symmetric domains. (This follows from the general theory, but since we have not defined Hermitian symmetric domains, we are merely stating this as a fact.)

Suppose we have an orthogonal direct sum
\[
V = \mathbb{R}^{a+2} = \mathbb{R}^a \oplus \mathbb{R}^2
\]
such that the pairings on \(\mathbb{R}^a\) and \(\mathbb{R}^2\) are defined by \(\begin{pmatrix} 1_{a-1} & -1 \end{pmatrix}\) and \(\begin{pmatrix} 1 & 1 \end{pmatrix} \), respectively, which have signatures \((a-1,1)\) and \((1,1)\) summing up to \((a,2)\) in total. Note that we have
\[
\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.
\]
Then we also have
\[
V_C \cong \mathbb{C}^a \oplus \mathbb{C}^2,
\]
and we denote the coordinates of \(\mathbb{C}^a\) and \(\mathbb{C}^2\) by \((z_1, \ldots, z_a)\) and \((w_1, w_2)\), respectively. With these coordinates, we have
\[
\tilde{H}_{SO,2} \cong \mathcal{H}_{SO,2} := \left\{ \begin{array}{ll}
z &\in \mathbb{C}^a : y = \text{Im}(z) > 0 \text{ in the sense that } \\
y \begin{pmatrix} 1_{a-1} & -1 \end{pmatrix} y < 0, \text{i.e., } y_2^2 > y_1^2 + \cdots + y_{a-1}^2 
\end{array} \right\}.
\]

The condition
\[
y_a^2 > y_1^2 + \cdots + y_{a-1}^2
\]
defines a familiar light cone with two connected components, one being defined by
\[
y_a > \sqrt{y_1^2 + \cdots + y_{a-1}^2},
\]
the other being defined by
\[
y_a < -\sqrt{y_1^2 + \cdots + y_{a-1}^2}.
\]
These two connected components correspond to two connected components \(H^+_{SO,2}\) and \(H^-_{SO,2}\) of \(H_{SO,2}\).
Note that
\[
\dim C \overline{H}_{SO_{a,2}} = \dim C H_{SO_{a,2}} = \dim C H_{SO_{a,2}}^+ = \dim C \mathbb{C} = a.
\]

Also, $H_{SO_{1,2}} \cong H_1^+$ and $H_{SO_{1,2}}^+ \cong H_1 = H$, when $a = 1$. In particular, we have obtained yet another generalization of the Poincaré upper half-plane $H$.

3.4.3. Connected components and spin groups. Under the assumption that $a \geq 1$, the real Lie group $SO_{a,2}(\mathbb{R})$ has two connected components (see, for example, [Hel01, Ch. X, Sec. 2, Lem. 2.4] or [Kna02, Prop. 1.145]), and the stabilizer of $H_{SO_{a,2}}$ in $G(\mathbb{R}) = SO_{a,2}(\mathbb{R})^+$ is the identity component $G(\mathbb{R})^+$; namely, the connected component containing the identity element (in the real analytic topology). The group $SO_{a,2}(\mathbb{R})$ can be viewed as the group of $\mathbb{R}$-points of the algebraic group $SO_{a,2}$ defined over $\mathbb{R}$, which admits a simply-connected two-fold covering group Spin$_{a,2}$. This is only a covering of algebraic groups, which means we can only expect a surjection on $\mathbb{C}$-points. In fact, the induced morphism
\[
(3.4.3.1) \quad \text{Spin}_{a,2}(\mathbb{R}) \to SO_{a,2}(\mathbb{R})
\]
on $\mathbb{R}$-points is not surjective, because Spin$_{a,2}(\mathbb{R})$ is connected and factors through $SO_{a,2}(\mathbb{R})^+$. Here we are keeping the language as elementary as possible, without introducing the spin groups in detail—see, for example, [GW09, Ch. 6], for a review of their construction based on Clifford algebras. But the connectedness of Spin$_{a,2}(\mathbb{R})$, nevertheless, is a general feature of the real points of any isotropic simply-connected connected simple algebraic group over $\mathbb{Q}$—see [PR94, Sec. 7.3, Thm. 7.6].

To understand this issue better, consider the $\text{Gal}(\mathbb{C}/\mathbb{R})$-equivariant short exact sequence
\[
(3.4.3.2) \quad 1 \to \{\pm 1\} \to \text{Spin}_{a,2}(\mathbb{C}) \to SO_{a,2}(\mathbb{C}) \to 1,
\]

which induces by taking Galois cohomology the long exact sequence
\[
(3.4.3.3) \quad 1 \to \{\pm 1\} \to \text{Spin}_{a,2}(\mathbb{R}) \to SO_{a,2}(\mathbb{R}) \to H^1(\text{Gal}(\mathbb{C}/\mathbb{R}), \{\pm 1\}) \cong \{\pm 1\},
\]

where the homomorphism
\[
(3.4.3.4) \quad SO_{a,2}(\mathbb{R}) \to \{\pm 1\}
\]

(induced by the connecting homomorphism) is the spinor norm. The kernel of the spinor norm is $SO_{a,2}(\mathbb{R})^+$, which is at the same time the image of Spin$_{a,2}(\mathbb{R})$.

We shall omit the verification that $SO_{a,2}(\mathbb{R})^+$ does act transitively on $H_{SO_{a,2}}^+$, with stabilizers isomorphic to
\[
(3.4.3.5) \quad SO_{a}(\mathbb{R}) \times SO_{2}(\mathbb{R}).
\]

Accordingly, $SO_{a,2}(\mathbb{R})$ acts transitively on $H_{SO_{a,2}}$, with stabilizers also isomorphic to $SO_{a}(\mathbb{R}) \times SO_{2}(\mathbb{R})$. We shall also omit the description of some bounded realizations $D_{SO_{a,2}}^+$ and $D_{SO_{a,2}}$ of $H_{SO_{a,2}}^+$ and $H_{SO_{a,2}}$, respectively.

In the classification in [Hel01, Ch. X, Sec. 6], $H_{SO_{a,2}} \cong D_{SO_{a,2}}^+$ (with $a \geq 1$) is a Hermitian symmetric domain of type BD I (with $(p,q) = (a,2)$ in the notation there). In Cartan’s classification, it is a bounded symmetric domain of type IV$_a$. 

3.4.4. Interpretation as conjugacy classes. Consider the homomorphism

\[(3.4.4.1) \quad h_0 : U_1 \to SO_{a,2}(\mathbb{R}) : e^{i\theta} \mapsto \begin{pmatrix} 1 & \cos 2\theta & -\sin 2\theta \\ 0 & \sin 2\theta & \cos 2\theta \end{pmatrix}.\]

Note that the image of \(h_0\) is connected because \(U_1\) is, and hence lies in \(SO_{a,2}(\mathbb{R})^+\). Moreover,

\[(3.4.4.2) \quad \text{Cent}_{h_0}(SO_{a,2}(\mathbb{R})^+) = \{ (A, D) \in SO_{a,2}(\mathbb{R})^+ \} \cong SO_a(\mathbb{R}) \times SO_2(\mathbb{R}) : (A, D) \mapsto (A, D).\]

Then \(K\) is a maximal compact subgroup of \(SO_{a,2}(\mathbb{R})^+\), and we have

\[(3.4.4.3) \quad H_{SO_{a,2}} = SO_{a,2}(\mathbb{R})^+ \cdot h_0 \cong SO_{a,2}(\mathbb{R})^+/ (SO_a(\mathbb{R}) \times SO_2(\mathbb{R})).\]

Note that there is a coefficient 2 of \(\theta\) in the above definition of \(h_0\). This \(h_0\) lifts to a homomorphism

\[(3.4.4.4) \quad \tilde{h}_0 : U_1 \to \text{Spin}_{a,2}(\mathbb{R}),\]

where \(\text{Spin}_{a,2}(\mathbb{R})\) is the above double cover of \(SO_{a,2}(\mathbb{R})^+\), and there is an injective homomorphism

\[(3.4.4.5) \quad \text{Spin}_{a,2}(\mathbb{R}) \hookrightarrow \text{Sp}_{2n}(\mathbb{R}),\]

with \(n = 2^a\) (cf. Examples 5.2.1.3 and 5.2.2.11 below), whose pre-composition with \(\tilde{h}_0\) lies in the conjugacy class of the analogous homomorphism for \(\text{Sp}_{2n}(\mathbb{R})\) in (3.1.4.1).

3.5. The case of \(E_7\). Our main references are [Bai70], [Kim93], [GL97], and [Eie09 Ch. 12].

3.5.1. Cayley numbers. Consider the algebra \(O\) of Cayley numbers, which is an octonion (i.e., degree eight) division algebra over \(\mathbb{R}\) which is normed, distributive, noncommutative, and nonassociative. We have

\[(3.5.1.1) \quad O = \mathbb{R} \oplus \mathbb{R}e_1 \oplus \cdots \oplus \mathbb{R}e_7\]

as an \(\mathbb{R}\)-vector space, which satisfies the multiplication rules given by

\[(3.5.1.2) \quad e_i^2 = -1\]

and

\[(3.5.1.3) \quad (e_ie_{i+1})e_{i+3} = e_i(e_{i+1}e_{i+3}) = 1\]

(with the indices periodically identified modulo 7) for all \(1 \leq i \leq 7\). We have

\[(3.5.1.4) \quad \mathcal{X} := x_0 - x_1e_1 - \cdots - x_7e_7\]

and

\[(3.5.1.5) \quad N(x) := x\mathcal{X} = x_0^2 + x_1^2 + \cdots + x_7^2 \geq 0\]

for each

\[(3.5.1.6) \quad x = x_0 + x_1e_1 + \cdots + x_7e_7 \in O,\]

where the square-root of \(N(x)\) defines the norm of \(x\).
3.5.2. An exceptional Jordan algebra over $\mathbb{R}$. Consider the Albert algebra
\begin{equation}
\mathfrak{A} := \text{Herm}_3(O),
\end{equation}
the $3 \times 3$ Hermitian matrices over $O$, which is a formally real Jordan algebra of exceptional type, with the Jordan algebra multiplication given by
\begin{equation}
X \cdot Y = \frac{1}{2}(XY + YX),
\end{equation}
where $XY$ and $YX$ are the usual matrix multiplications. (See, for example, [FK95, Ch. II–V and VIII] for a general introduction to Jordan algebras and the classification of formally real ones. See also [AMRT75, Ch. II, Rem. at the end of Sec. 1] or [AMRT10, Ch. II, Rem. 1.11], and the sections following there, for why they matter.)

For
\begin{equation}
X = \begin{pmatrix}
a & x & y \\
b & b & z \\
n & y & c
\end{pmatrix} \in \mathfrak{A}
\end{equation}
(with $a, b, c \in \mathbb{R}$ and $x, y, z \in O$), we can define the trace
\begin{equation}
\text{tr}(X) = a + b + c
\end{equation}
and the determinant
\begin{equation}
\text{det}(X) = abc - aN(z) - bN(y) - cN(x) + \text{tr}((xz)y).
\end{equation}
Then we also have an inner product
\begin{equation}
(X, Y) := \text{tr}(X \cdot Y)
\end{equation}
and a trilinear form $(\cdot, \cdot, \cdot)$ satisfying the identity
\begin{equation}
(X, X, X) = \text{det}(X).
\end{equation}
This trilinear form defines a cross product $X \times Y$ such that
\begin{equation}
(X \times Y, Z) = 3(X, Y, Z),
\end{equation}
for $X, Y, Z \in \mathfrak{A}$, which specializes to
\begin{equation}
X \times X = \begin{pmatrix}
bc - N(z) & yz - cx & xz - by \\
z\overline{y} - cx & ac - N(y) & xy - az \\
\overline{x}y - by & \overline{y}x - az & ab - N(x)
\end{pmatrix}
\end{equation}
when $X = Y$.

Since $O$ is nonassociative, it is rather a miracle that the above definitions can be made for $3 \times 3$ Hermitian matrices over $O$—nothing similar works for larger matrices.

3.5.3. A real Lie group with Lie algebra $\mathfrak{e}_{7(-25)}$. Consider the Lie group
\begin{equation}
M := \{(g, r) \in \text{GL}(\mathfrak{A}) \times \mathbb{R}^\times : \text{det}(gX) = r \text{det}(X)\},
\end{equation}
so that we have a homomorphism
\begin{equation}
\nu : M \rightarrow \mathbb{R}^\times : (g, r) \mapsto r.
\end{equation}
Then we have
\begin{equation}
\ker(\nu) = \{g \in \text{GL}(\mathfrak{A}) : \text{det}(gX) = \text{det}(X)\}.
\end{equation}
It is known that $\ker(\nu)$ is a connected semisimple real Lie group of real rank 2 with Lie algebra isomorphic to $\mathfrak{e}_{6(-26)}$. The number $-26$ in the notation of $\mathfrak{e}_{6(-26)}$
means it is the real form of the complex simple Lie algebra of type $E_6$ such that the Killing form has signature $(a, b)$ with $a - b = -26$; or, equivalent, that every Cartan involution has trace $26 = -(-26)$. For complex simple Lie algebras of exceptional type, such numbers already classify their real forms up to isomorphism. (The same explanation applies to other real Lie algebras of exceptional type below.) Also, the center of $M$ is

\[(3.5.3.4) \quad Z(M) \cong \mathbb{R}^x\]

with $x \in \mathbb{R}^x$ acting as multiplication by the scalar $x$ on $\mathfrak{a}$, so that $\nu(x) = x^3$. In this case, the Lie algebra of $M$ is the direct sum of those of $Z(M)$ and $\ker(\nu)$, and $\ker(\nu)$ is the derived group of $M$. The $\ker(\nu)$-orbit of 1 in $\mathfrak{a}$ is isomorphic to the quotient of $\ker(\nu)$ by a maximal compact subgroup with Lie algebra isomorphic to $f_4$. In the classification in [Hel01, Ch. X, Sec. 6], this orbit is a noncompact Riemannian symmetric space of type $E IV$, which is not Hermitian.

Moreover, $g \in M$ induces $g^* \in M$ by

\[(3.5.3.5) \quad (gX, g^*Y) = (X, Y).\]

Then $M$ acts on

\[(3.5.3.6) \quad \mathfrak{m} := \mathfrak{a} \oplus \mathbb{R} \oplus \mathfrak{a} \oplus \mathbb{R}\]

by

\[(3.5.3.7) \quad g(X, \varepsilon, X', \varepsilon') := (gX, \nu(g)^{-1}\varepsilon, g^*X', \nu(g)\varepsilon').\]

There is also an additive action

\[(3.5.3.8) \quad \rho : \mathfrak{a} \to \text{GL}(\mathfrak{m}) : B \mapsto \rho_B\]

with $\rho_B$ defined by

\[(3.5.3.9) \quad \rho_B \left( \begin{array}{c} X \\ \varepsilon \\ X' \\ \varepsilon' \end{array} \right) := \left( \begin{array}{c} X + 2B \times X' + \varepsilon B \times B \\ \varepsilon \\ X' + \varepsilon B \\ \varepsilon' + (B, X) + (B \times X', \varepsilon) + \varepsilon \det(B) \end{array} \right),\]

with elements of $\mathfrak{m}$ represented by column vectors (cf. (3.5.3.6)).

Consider the Lie group

\[(3.5.3.10) \quad U := \text{image}(\rho).\]

It is known that $M$ normalizes $U$. Let $\iota \in \text{GL}(\mathfrak{m})$ be defined by

\[(3.5.3.11) \quad \iota(X, \varepsilon, X', \varepsilon') := (-X', -\varepsilon', X, \varepsilon).\]

Consider the Lie groups

\[(3.5.3.12) \quad P := M \ltimes U\]

and

\[(3.5.3.13) \quad G := \langle \iota, P \rangle\]

(i.e., $G$ is generated by $\iota$ and $P$) as subgroups of $\text{GL}(\mathfrak{m})$. It is known that $G$ is a connected semisimple Lie group of real rank 3 with Lie algebra isomorphic to $e_7(-25)$, and that $P$ is the maximal parabolic subgroup of $G$ stabilizing the subspace \{(0, 0, 0, \varepsilon') : \varepsilon' \in \mathbb{R} \} of \mathfrak{m}.

These definitions might seem rather complicated and unmotivated, but it is worth drawing the analogy of the element $\iota$ and the subgroups $U$, $M$, and $P$ of
G with the corresponding element $\begin{pmatrix} -1_n & 1_n \\ \ast & \ast \end{pmatrix}$ and the subgroups $\left\{ \begin{pmatrix} 1_n & \ast \\ \ast & 1_n \end{pmatrix} \right\}$, $\left\{ \begin{pmatrix} \ast & \ast \\ \ast & \ast \end{pmatrix} \right\}$, and $\left\{ \begin{pmatrix} \ast & \ast \\ \ast & \ast \end{pmatrix} \right\}$ of $\text{Sp}_{2n}(\mathbb{R})$.

A more systematic definition is that $G$ is the Lie subgroup of $\text{GL}(W)$ consisting of elements preserving both the quartic form

\begin{equation}
Q(X, \varepsilon, X', \varepsilon') := (X \times X, X' \times X') - \varepsilon \det(X) - \varepsilon' \det(X') - \frac{1}{4}((X, X') - \varepsilon \varepsilon')^2
\end{equation}

(3.5.3.14)

and the skew-symmetric bilinear form

\begin{equation}
\langle (X_1, \varepsilon_1, X'_1, \varepsilon'_1), (X_2, \varepsilon_2, X'_2, \varepsilon'_2) \rangle := (X_1, X'_2) - (X_2, X'_1) + \varepsilon_1 \varepsilon'_2 - \varepsilon_2 \varepsilon'_1
\end{equation}

on $W$. But, as we shall see, it is useful to know the explicit facts (3.5.3.13) and (3.5.3.12).

3.5.4. An octonion upper half-space. Consider

\begin{equation}
\mathcal{H}_{E_7} := \{ Z \in \mathfrak{A} \otimes \mathbb{C} : \text{Im}(Z) > 0 \},
\end{equation}

where $\text{Im}(Z) > 0$ means the left multiplication of $\text{Im}(Z)$ on $\mathfrak{A}$ induces a positive definite linear transformation, or equivalently that $\text{Im}(Z)$ lies in the interior of the set $\{ X^2 : X \in \mathfrak{A} \}$ with its natural real analytic topology (see [FK95, Thm. III.2.1]). (For an analogy, consider the case of symmetric matrices over $\mathbb{R}$, where positive definite matrices are exactly those matrices in the interior of the set of squares of symmetric matrices with its natural real analytic topology.) This is the most exotic generalization of the various classical upper half-spaces in Sections 3.1.1, 3.2.3, and 3.3.2 defined by Hermitian matrices over $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$, respectively. Together with the “semi-classical” examples in Section 3.4.2 these are all the irreducible examples (namely, ones that are not isomorphic to nontrivial products of smaller ones) of the so-called tube domains. They are all of the form

\begin{equation}
\{ X + iY : X, Y \in \mathfrak{J}, Y > 0 \} \subset \mathfrak{J} \otimes \mathbb{C},
\end{equation}

where $\mathfrak{J}$ is a formally real Jordan algebra (see [FK95, Ch. VIII–X]; see also [AMRT75, Ch. III, Sec. 1] or [AMRT10, Ch. III, Sec. 1]), where the symbol $i$ denotes the action of $1 \otimes i$ on $\mathfrak{J} \otimes \mathbb{C}$. In the definition of $\mathcal{H}_{E_7}$, we have $\mathfrak{J} = \mathfrak{A} = \text{Herm}_3(\mathbb{O})$.

Note that

\begin{equation}
\dim_{\mathbb{C}} \mathcal{H}_{E_7} = \dim_{\mathbb{C}}(\mathfrak{A} \otimes \mathbb{C}) = \dim_{\mathbb{R}}(\mathfrak{A}) = 27.
\end{equation}

The group $U$ acts on $\mathcal{H}_{E_7}$ by the translation

\begin{equation}
Z \mapsto Z + B,
\end{equation}

for each $\rho_B \in U$ associated with $B \in \mathfrak{A}$. The group $M$ acts on $\mathcal{H}_{E_7}$ by its natural action on $\mathfrak{A} \otimes \mathbb{C}$ (up to a sign convention), which preserves the subset $\mathcal{H}_{E_7}$. The element $i \in G$ acts on $\mathcal{H}_{E_7}$ by

\begin{equation}
i(Z) := -Z^{-1},
\end{equation}

where $i$ is the antiholomorphic involution.
where $Z^{-1}$ is defined by inverting the Jordan algebra multiplication in $\mathbb{A} \otimes \mathbb{C}$. One can check that these actions are compatible with each other, and define an action of $G$ on $\mathcal{H}_{E_7}$.

Given any $Z = X + iY \in \mathcal{H}_{E_7}$, where $X, Y \in \mathbb{A}$ and where the symbol $i$ denotes the action of $1 \otimes i$ on $\mathbb{A} \otimes \mathbb{C}$, we have

$$\rho_X(Z) = iY$$

for some $Y > 0$. By Freudenthal’s theorem (which intriguingly asserts that matrices in $\text{Herm}_3(\mathbb{O})$ can also be diagonalized; see [FK95, Thm. V.2.5]), there exists some $g \in M$ such that

$$gY = 1_3,$$

and so that

$$g(iY) = i1_3.$$

This shows that

$$\mathcal{H}_{E_7} = G \cdot i1_3$$

and so that the action of $G$ on $\mathcal{H}_{E_7}$ is transitive.

Let $K$ denote the stabilizer of $i1_3$ in $G$. Then

$$K = \text{Cent}_i(G)$$

as subgroups of $G$, which is a maximal compact subgroup of $G$ and is a connected compact Lie group with Lie algebra $\mathfrak{e}_6 \oplus \mathbb{R}$. Consider the homomorphism

$$h_0 : U_1 \rightarrow G$$

defined in block matrix form by setting

$$h_0(\cos \theta + i \sin \theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \in \text{GL}(\mathfrak{m})$$

(by left multiplication on column vectors representing elements in $\mathfrak{m}$; cf. (3.5.3.6)). In particular, $h_0(i) = \iota$, so that we also have

$$\mathcal{H}_{E_7} = G \cdot h_0 \cong G/K.$$

There is no homomorphism of the form $G \rightarrow \text{Sp}_{2n}(\mathbb{R})$, for any $n \geq 1$, whose pre-composition with $h_0 : U_1 \rightarrow G$ lies in the conjugacy class of the analogous homomorphism for $\text{Sp}_{2n}(\mathbb{R})$ in (3.1.4.1) (see [Del79] 1.3.10(i)) or [MIL13, Sec. 10, p. 531]; cf. the end of Section 5.2.2 below).

3.5.5. Bounded realization. There is also a bounded realization $D_{E_7}$ of $\mathcal{H}_{E_7}$, also with coordinates in $\mathbb{A} \otimes \mathbb{C}$, with a change of coordinates

$$\mathcal{H}_{E_7} \overset{\sim}{\rightarrow} D_{E_7} : Z \mapsto (Z - i1_3)(Z + i1_3)^{-1},$$

where $(Z + i1_3)^{-1}$ is (as before) defined by inverting the Jordan algebra multiplication in $\mathbb{A} \otimes \mathbb{C}$, which is yet another generalization of the classical Cayley transform (cf. Sections 3.1.3 and 5.2.4).
In the classification in [Hel01, Ch. X, Sec. 6], $H_{E_7} \cong D_{E_7}$ is a Hermitian symmetric domain of type E VII. In Cartan’s classification, it is a bounded symmetric domain of type VI.


3.6.1. Bounded and unbounded realizations. There is (up to isomorphism) just one irreducible Hermitian symmetric domain that we have not discussed yet. (Again, being irreducible means not isomorphic to a nontrivial products of smaller ones.) Let us start with a bounded realization, given by

\[
D_{E_6} := D_{E_7} \cap \left\{ \begin{pmatrix} 0 & x & y \\ x & 0 & 0 \\ y & 0 & 0 \end{pmatrix} : x, y \in O \otimes \mathbb{C} \right\}
\]

in $A \otimes \mathbb{C}$ (see Section 3.5.5), where $\overline{x}$ and $\overline{y}$ are the conjugations induced by the ones in $O$, which can be embedded as an open submanifold of $O^2 \otimes \mathbb{C}$. Let us also let $H_{E_6}$ denote the preimage of $D_{E_6}$ under the isomorphism (3.5.5.1), which is then an unbounded realization.

Essentially by construction, we have

\[
\dim_{\mathbb{C}} H_{E_6} = \dim_{\mathbb{C}} D_{E_6} = 2 \dim_{\mathbb{C}} (O \otimes \mathbb{C}) = 2 \dim_{\mathbb{R}} O = 16.
\]

By [Hir66, Sec. 3], $D_{E_6}$ and hence $H_{E_6}$ admit transitive actions of a connected semisimple Lie group $G$ of rank 2 with Lie algebra isomorphic to $\mathfrak{e}_{6(-14)}$, with stabilizers given by maximal compact subgroups of $G$ which are connected Lie groups with Lie algebras isomorphic to $\mathfrak{so}_{10} \oplus \mathbb{R}$.

In the classification in [Hel01, Ch. X, Sec. 6], $H_{E_6} \cong D_{E_6}$ is a Hermitian symmetric domain of type E III. In Cartan’s classification, it is a bounded symmetric domain of type V.

3.6.2. A real Lie group of Lie algebra $\mathfrak{e}_{6(-14)}$. It is not easy to describe $D_{E_6}$ and $H_{E_6}$, together with the actions of certain $G$ as above, in explicit coordinates. Nevertheless, let us define a real Lie group $G$ with Lie algebra $\mathfrak{e}_{6(-14)}$, which can be given the structure of the $\mathbb{R}$-points of an algebraic group over $\mathbb{Q}$, and which acts on $D_{E_6}$ and $H_{E_6}$ by the abstract theory of Hermitian symmetric domains.

Consider the pairing

\[
\langle X, Y \rangle := \langle X^c, Y \rangle,
\]

for $X, Y \in A \otimes \mathbb{C}$, where $c$ denotes the complex conjugation with respect to the tensor factor $\mathbb{C}$; consider the involution $\sigma$ on $A = \text{Herm}_3(O)$ (which extends linearly to $A \otimes \mathbb{C}$) defined by

\[
\begin{pmatrix} a & x & y \\ \overline{x} & b & z \\ \overline{y} & \overline{z} & c \end{pmatrix}^\sigma = \begin{pmatrix} a & -x & -y \\ -\overline{x} & b & z \\ -\overline{y} & \overline{z} & c \end{pmatrix};
\]

and consider

\[
G := \{ g \in \text{GL}(A \otimes \mathbb{C}) : \det(gX) = \det(X), \langle gX, gY \rangle_\sigma = \langle X, Y \rangle_\sigma \},
\]

where

\[
\langle X, Y \rangle_\sigma := \langle X^\sigma, Y \rangle.
\]
We note that
\[(3.6.2.5) \quad \{ g \in \text{GL}(\mathfrak{A} \otimes \mathbb{C}) : \det(gX) = \det(X) \}\]
is a complex Lie group with Lie algebra $\mathfrak{e}_6(\mathbb{C})$, and so the upshot is that the condition
\[(3.6.2.6) \quad \langle gX, gY \rangle_\sigma = \langle X, Y \rangle_\sigma\]
defines a real form of rank 2 of this complex Lie group with Lie algebra isomorphic to $\mathfrak{e}_6(-14)$, which is not isomorphic to either its compact form (see \[(3.5.4.10)\] and the remark following it) or the noncompact real form
\[(3.6.2.7) \quad \{ g \in \text{GL}(\mathfrak{A}) : \det(gX) = \det(X) \}\]
(see \[(3.5.3.3)\] and the remark following it) whose Lie algebra is isomorphic to $\mathfrak{e}_6(-26)$.

Thus far, we have seen three kinds of real Lie groups of type $E_6$ in our examples, with one compact and the other two of real rank 2.

3.6.3. Interpretation as conjugacy classes. Consider the homomorphism
\[(3.6.3.1) \quad h_0 : U_1 \to G\]
defined by
\[(3.6.3.2) \quad h_0(\alpha) \begin{pmatrix} a & x & y \\ \bar{y} & b & z \\ y & z & c \end{pmatrix} = \begin{pmatrix} \alpha^4 a & \alpha x & \alpha y \\ \alpha \bar{x} & \alpha^{-2} b & \alpha^{-2} z \\ \alpha \bar{y} & \alpha^{-2} \bar{x} & \alpha^{-2} c \end{pmatrix},\]
with $\alpha$ acting as $1 \otimes \alpha$ on $\mathfrak{O} \otimes \mathbb{C}$. Then it can be checked directly that
\[(3.6.3.3) \quad K := \text{Cent}_{h_0}(G),\]
is a maximal compact subgroup of $G$, which is a connected Lie group with Lie algebra isomorphic to $\mathfrak{so}_{10} \oplus \mathbb{R}$, so that we have
\[(3.6.3.4) \quad \mathcal{H}_{E_6} = G \cdot h_0 \cong G/K.\]

There is no homomorphism of the form $G \to \text{Sp}_{2n}(\mathbb{R})$, for some $n \geq 1$, whose pre-composition with $h_0 : U_1 \to G$ lies in the conjugacy class of the analogous homomorphism for $\text{Sp}_{2n}(\mathbb{R})$ in \[(3.1.4.1)\] (see \cite[1.3.10(i)]{Del79} or \cite[Sec. 10, p. 530]{Mil13}; cf. the end of Section 5.2.2 below).

3.7. A brief summary. Let us summarize the groups and symmetric domains we have considered as follows. Let us also include the information of the corresponding special node on each Dynkin diagram, marked with a unique double circle on the diagram. Such special nodes have not been explained, but should be useful for the readers when they consult more advanced texts.

1. $G(\mathbb{R}) = \text{Sp}_{2n}(\mathbb{R}), \ K \cong U_n, \ D = \mathcal{H}_n$:
\[(3.7.1) \quad \bullet \quad \circ \cdots \circ \quad \bullet \]

2. $G(\mathbb{R}) = U_{a,b}, \ K \cong U_a \times U_b, \ D = \mathcal{H}_{a,b}$:
\[(3.7.2) \quad \bullet \quad \circ \cdots \circ \quad \bullet \]

Here the special node is the $b$-th from the left, or alternatively, from the right, because it is only up to automorphisms of the Dynkin diagram.
(3) \( G(\mathbb{R}) = \text{SO}^+_n, \ K \cong U_n, \ D = \mathcal{H}_{\text{SO}^+_n}, \) with \( n \geq 2: \)

\[
\begin{array}{c}
\circ - \circ - \circ - \cdots \\
\end{array}
\]

(3.7.3)

Again, the special node is only up to automorphisms of the Dynkin diagram. Therefore, it can be, alternatively, the other right-most node.

(4) \( G(\mathbb{R})^+ = \text{SO}_{a,2}(\mathbb{R})^+, \ K \cong \text{SO}_a(\mathbb{R}) \times \text{SO}_2(\mathbb{R}), \ D^+ = \mathcal{H}^+_{\text{SO}_{a,2}}, \) with \( a \geq 1 \) but \( a \neq 2: \)

(a) \( a = 2n - 1 \) is odd, with \( n \geq 1: \)

\[
\begin{array}{c}
\circ - \circ - \cdots \\
\end{array}
\]

(3.7.4)

(b) \( a = 2n - 2 \) is even, with \( n \geq 3: \)

\[
\begin{array}{c}
\circ - \circ - \cdots \\
\end{array}
\]

(3.7.5)

(In the excluded case \( a = 2, \) we have \( \text{Spin}_{2,2}(\mathbb{R}) \cong \text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R}), \) and \( \mathcal{H}^+_{\text{SO}_{2,2}} \cong \mathcal{H} \times \mathcal{H} \) is not irreducible.)

(5) \( \text{Lie } G \cong \mathfrak{e}_{6(-14)}, \ \text{Lie } K \cong \mathfrak{so}_{10} \oplus \mathbb{R}, \ D = \mathcal{H}_{E_6}: \)

\[
\begin{array}{c}
\circ \\
\end{array}
\]

(3.7.6)

Yet again, the special node is only up to automorphisms of the Dynkin diagram. Therefore, it can be, alternatively, the right-most node.

(6) \( \text{Lie } G \cong \mathfrak{e}_{7(-25)}, \ \text{Lie } K \cong \mathfrak{e}_6 \oplus \mathbb{R}, \ D = \mathcal{H}_{E_7}: \)

\[
\begin{array}{c}
\circ \\
\end{array}
\]

(3.7.7)

These exhaust all the possibilities of irreducible Hermitian symmetric domains.

In the cases \( \lbrack 1 \rbrack, \ \lbrack 2 \rbrack, \) and \( \lbrack 3 \rbrack, \) there exist injective homomorphisms of the form \( G(\mathbb{R}) \to \text{Sp}_{2n'}(\mathbb{R}), \) for some \( n' \geq 1, \) whose pre-composition with \( h_0 : U_1 \to G(\mathbb{R}) \) lies in the conjugacy class of the analogous homomorphism for \( \text{Sp}_{2n'}(\mathbb{R}) \) as in (3.1.4.1). In the case \( \lbrack 4 \rbrack, \) the analogous assertion is true after replacing \( \text{SO}_{a,2}(\mathbb{R})^+ \) with its double cover \( \text{Spin}_{a,2}(\mathbb{R}). \) In the cases \( \lbrack 5 \rbrack \) and \( \lbrack 6 \rbrack, \) the analogous assertion is simply false. (In Section 5.2.2 below, we will encounter this issue again in the classification of abelian-type Shimura data, although the question there is more subtle because we need to also consider the structure of algebraic groups over \( \mathbb{Q}. \))

4. Rational boundary components and compactifications

In this section, we will explain how the double coset spaces

\[
\mathcal{X}_U = G(\mathbb{Q}) \backslash (D \times G(\mathbb{A}^\infty))/U \cong \prod_{\iota \in I} \Gamma_\iota \backslash \mathcal{D}^+ 
\]

as in \( \lbrack 2.2.15 \rbrack \) and \( \lbrack 2.2.21 \rbrack \) are compactified and shown to be algebraic varieties in cases where \( D \) are Hermitian symmetric domains or their finite unions.
4.1. Examples of rational boundary components. Let us start by explaining the notion of rational boundary components in many examples, based on the Hermitian symmetric domains we have seen in Section 3. We will begin with the cases with \( \mathcal{D} = \mathcal{H} \) in Section 4.1.1, followed by the cases with \( \mathcal{D} = \mathcal{H} \times \mathcal{H} \) in Section 4.1.2. We have to consider quotients of the form \( \Gamma \mathcal{D} \).

Cases with \( \mathcal{H} \) in Section 4.1.1, followed by the cases with \( \mathcal{D} = \mathcal{H} \times \mathcal{H} \) in Section 4.1.2; the cases with \( \mathcal{D} = \mathcal{H}_{a,b} \), with \( a \geq b \), in Section 4.1.3. We will begin with the cases in Section 4.1.7; and the cases with \( \mathcal{D} = \mathcal{H}_{E_6} \) or \( \mathcal{D} = \mathcal{H}_{E_7} \) in Section 4.1.7.

4.1.1. Cases with \( \mathcal{D} = \mathcal{H} \). Suppose \( G = \text{SL}_2 \), so that we have \( \mathcal{D} = \mathcal{H} \) as in Examples 2.2.7 and 2.2.23 and we have to consider quotients of the form \( \Gamma \mathcal{D} \), where \( \Gamma \) is an arithmetic subgroup of \( \text{SL}_2(\mathbb{Q}) \). Note that the action of \( \text{SL}_2(\mathbb{Q}) \) preserves

\[
\mathcal{H}^* := \mathcal{H} \cup \mathbb{P}^1(\mathbb{Q})
\]

as a subset of \( \mathbb{P}^1(\mathbb{C}) \).

In the natural topology of \( \mathbb{P}^1(\mathbb{C}) \), the boundary of \( \mathcal{H} \) is \( \mathbb{P}^1(\mathbb{R}) = \text{SL}_2(\mathbb{R}) \cdot \infty \), where \( \infty \) means just \( \left( \begin{array}{c} 1 \\ 0 \end{array} \right) \) in projective coordinates. The reason to consider instead the rational subset \( \mathbb{P}^1(\mathbb{Q}) = \text{SL}_2(\mathbb{Q}) \cdot \infty \) is the reduction theory: The fundamental domain of \( \mathcal{H} \) under the action of \( \Gamma \) can be compactified by adding only points in \( \mathbb{P}^1(\mathbb{Q}) = \text{SL}_2(\mathbb{Q}) \cdot \infty \). The irrational points are completely irrelevant. Therefore, by topologizing \( \mathcal{H}^* \) at the points of \( \mathbb{P}^1(\mathbb{Q}) \) using open discs bounded by the so-called horocycles, namely, circles tangent to \( \mathbb{P}^1(\mathbb{R}) \) at the points of \( \mathbb{P}^1(\mathbb{Q}) \) (see [Shi71, Sec. 1.3]), we obtain a compactification

\[
\Gamma \backslash \mathcal{H} \rightarrow \Gamma \backslash \mathcal{H}^*
\]

such that \( \Gamma \backslash \mathcal{H}^* \) has the structure of a compact Riemann surface (or, in other words, a complete algebraic curve over \( \mathbb{C} \)), with finitely many boundary points

\[
\Gamma \backslash \mathcal{H}^* - \Gamma \backslash \mathcal{H} \cong \Gamma \backslash \mathbb{P}^1(\mathbb{Q})
\]

called cusps. (Here the finiteness of \( \Gamma \backslash \mathbb{P}^1(\mathbb{Q}) \) can be deduced from the elementary fact that \( \mathbb{P}^1(\mathbb{Q}) = \text{SL}_2(\mathbb{Q}) \cdot \infty = \text{SL}_2(\mathbb{Z}) \cdot \infty \). The points of \( \mathbb{P}^1(\mathbb{Q}) = \text{SL}_2(\mathbb{Q}) \cdot \infty \) are the so-called rational boundary components of \( \mathcal{H} \), which are prototypical examples of their analogues for more general Hermitian symmetric domains.

Note that the notions of cusps and rational boundary components depend on the structure of \( G = \text{SL}_2 \) as an algebraic group over \( \mathbb{Q} \) in an essential way. There exist other algebraic groups \( G' \) over \( \mathbb{Q} \) such that \( G'(\mathbb{R}) \cong G(\mathbb{R}) = \text{SL}_2(\mathbb{R}) \) and such that the fundamental domains for the actions of the corresponding arithmetic subgroups \( \Gamma' \) of \( G'(\mathbb{Q}) \) on \( \mathcal{D} \) are already compact, in which case there is no need to add any cusps (cf. the example of Shimura curves in Section 6.2.2 below).

4.1.2. Cases with \( \mathcal{D} = \mathcal{H} \times \mathcal{H} \). There are two important cases of \( G \) with \( G(\mathbb{R}) \) acting on \( \mathcal{D} = \mathcal{H} \times \mathcal{H} \).

The first \( G \) is simply \( \text{SL}_2 \times \text{SL}_2 \). Then we should take

\[
\mathcal{D}^* := \mathcal{H}^* \times \mathcal{H}^*
\]

\[= (\mathcal{H} \times \mathcal{H}) \cup (G(\mathbb{Q}) \cdot (\mathcal{H} \times \infty)) \cup (G(\mathbb{Q}) \cdot (\infty \times \mathcal{H})) \cup (G(\mathbb{Q}) \cdot (\infty \times \infty))\]

as subsets of \( \mathbb{P}^1(\mathbb{C}) \times \mathbb{P}^1(\mathbb{C}) \). There are three \( G(\mathbb{Q}) \)-orbits of rational boundary components (which are those orbits other than \( \mathcal{H} \times \mathcal{H} \)).
The other $G$ is the restriction of scalars $\text{Res}_{F/\mathbb{Q}} \text{SL}_2$, for some real quadratic extension field $F$ of $\mathbb{Q}$. It is the group scheme $G$ over $\mathbb{Q}$ whose $R$-points are

$$G(R) = \text{SL}_2(R \otimes \mathbb{Q}),$$

for each $\mathbb{Q}$-algebra $R$. In particular, we have $G(\mathbb{Q}) = \text{SL}_2(F)$. This is the case of Hilbert modular surfaces, which are generalized to higher-dimensional Hilbert modular varieties by replacing $F$ with totally real extension fields of $\mathbb{Q}$ of higher degrees. Then we should take

$$D^* := (H \times H) \cup (G(\mathbb{Q}) \cdot (\infty \times \infty)),$$

with only one $G(\mathbb{Q})$-orbit of rational boundary components (which is the unique orbit other than $H \times H$).

In both cases, we can topologize $D^*$ such that $\Gamma \backslash D^*$ is a compactification of $\Gamma \backslash D$ for each arithmetic subgroup $\Gamma$ of $G(\mathbb{Q})$. When $G$ is simply $\text{SL}_2 \times \text{SL}_2$, this is just doubling the case where $G = \text{SL}_2$. When $G$ is $\text{Res}_{F/\mathbb{Q}} \text{SL}_2$, this is essentially proved in [Shi63a], in a different language. This generalizes the case where $G = \text{SL}_2$ but will be superseded by the introduction of Satake topology in general.

Here are some general principles, which are valid for not only these two examples, but also for the forthcoming ones:

1. The rational boundary components are in one-one correspondence with (rational) parabolic subgroups $P$ of the algebraic group $G$ over $\mathbb{Q}$ whose pullback to every $\mathbb{Q}$-simple (almost) factor $G'$ of $G$ is either the whole $G'$ or a maximal proper parabolic subgroup $P'$. When the group $G$ is $\mathbb{Q}$-simple, we will just say that such a $P$ is a maximal parabolic subgroup.

2. For classical groups $G$ defined by nondegenerate pairings over finite-dimensional semisimple algebras over $\mathbb{Q}$, the (rational) parabolic subgroups of $G$ are always stabilizers of increasing sequences of totally isotropic subspaces.

4.1.3. Cases with $D = H_n$. Suppose $G = \text{Sp}_{2n}$, so that $G(\mathbb{R})$ acts on $H_n$ as in Section 3.1.1.

When $n = 2$, there are three $G(\mathbb{Q})$-orbits of proper parabolic subgroups of $G(\mathbb{Q})$, with the following representatives, respectively: (We warn the readers that, although we will write down only the groups of $\mathbb{Q}$-points, we will still be working with algebraic groups over $\mathbb{Q}$. This is a somewhat misleading abuse of language, which we adopt only for simplicity of exposition.)

1. A Borel subgroup

$$B(\mathbb{Q}) = \left\{ \begin{pmatrix} * & * & * \\ * & * & * \\ * & * \end{pmatrix} \right\}$$

(which is a minimal parabolic subgroup) stabilizing the increasing sequence

$$\left\{ \begin{pmatrix} * \\ 0 \\ 0 \end{pmatrix} \right\} < \left\{ \begin{pmatrix} * \\ 0 \end{pmatrix} \right\}.$$
of totally isotropic subspaces of $\mathbb{Q}^4$. Note that we have

$$\begin{align*}
\left\{ \begin{pmatrix} * & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \right\} \perp \left\{ \begin{pmatrix} * \\ * \\ 0 \\ 0 \end{pmatrix} \right\}
\end{align*}$$

and

$$\begin{align*}
\left\{ \begin{pmatrix} * \\ * \\ 0 \\ 0 \end{pmatrix} \right\} \perp \left\{ \begin{pmatrix} * \\ * \\ 0 \\ 0 \end{pmatrix} \right\}
\end{align*}$$

with respect to the alternating pairing

$$\langle \cdot, \cdot \rangle : \mathbb{Q}^4 \times \mathbb{Q}^4 \to \mathbb{Q}$$

defined by $J_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and so we can complete the above sequence (4.1.3.2) into a flag

$$\begin{align*}
\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\} \subset \left\{ \begin{pmatrix} * \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\} \subset \left\{ \begin{pmatrix} * \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\} \subset \left\{ \begin{pmatrix} * \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\} \subset \left\{ \begin{pmatrix} * \\ * \\ * \\ * \end{pmatrix} \right\}
\end{align*}$$

in $\mathbb{Q}^4$. This also explains why the entries of $B$ are not exactly upper-triangular. Sometimes it is convenient to work with upper-triangular matrices—then one should replace $J_2$ with a matrix that is indeed anti-diagonal, not just block anti-diagonal, as remarked in Example 2.1.2.

(2) The **Klingen parabolic subgroup**

$$P^{(1)}(\mathbb{Q}) = \left\{ \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \right\}$$

(which is a maximal parabolic subgroup) stabilizing the totally isotropic subspace

$$\left\{ \begin{pmatrix} * \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\}$$

of $\mathbb{Q}^4$.

(3) The **Siegel parabolic subgroup**

$$P^{(2)}(\mathbb{Q}) = \left\{ \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \right\}$$
(which is a maximal parabolic subgroup) stabilizing the totally isotropic subspace

\[
\begin{pmatrix}
  * & * \\
  0 & 0 \\
\end{pmatrix}
\]

of \( \mathbb{Q}^4 \).

The Borel subgroup B is not maximal, and hence will not be needed in the consideration of rational boundary components. Nevertheless, we have

\[
B(\mathbb{Q}) = P^{(1)}(\mathbb{Q}) \cap P^{(2)}(\mathbb{Q})
\]

by definition, and \( P^{(1)}(\mathbb{Q}) \) and \( P^{(2)}(\mathbb{Q}) \) will both be needed.

Now that we have seen the maximal parabolic subgroups when \( n = 2 \), let us consider a more general \( n \). The rational boundary components of \( H_n \) are given by elements of the \( G(\mathbb{Q}) \)-orbits of

\[
(4.1.3.12) \quad \begin{pmatrix} \infty_r & \cdot \\
 H_{n-r} & \cdot \\
\end{pmatrix} \cong \mathcal{H}_{n-r},
\]

which means

\[
(4.1.3.13) \quad \left\{ \begin{pmatrix} 1_r & Z \\
 0_r & 1_{n-r} \\
\end{pmatrix} : Z \in \mathcal{H}_{n-r} \right\}
\]

in generalized projective coordinates (cf. Remark 3.1.1.4), for \( r = 1, \ldots, n \). (When \( r = 0 \), we just get \( H_n \) itself.)

There is a topology, called the Satake topology (see [Sat60], [BB66, Thm. 4.9], and [BJ06, Sec. III.3] for systematic treatments not just for this case), on the set

\[
(4.1.3.14) \quad \mathcal{H}_n^* := H_n \cup \left( \bigcup_{1 \leq r \leq n} G(\mathbb{Q}) \cdot \mathcal{H}_{n-r} \right),
\]

where \( \mathcal{H}_{n-r} \) stands for \( \begin{pmatrix} \infty_r & \cdot \\
 H_{n-r} & \cdot \\
\end{pmatrix} \) as in (4.1.3.12), such that \( \mathcal{H}_{n-r} \) lies in the closure of \( \mathcal{H}_{n-s} \) exactly when \( r \geq s \), and such that \( \Gamma \backslash \mathcal{H}_n^* \) is a compactification of \( \Gamma \backslash H_n \) as a topological space, for every arithmetic subgroup \( \Gamma \) of \( G(\mathbb{Q}) \) (by the reduction theory, as in [BHC62] and [Bor62]). According to [BB66], \( \Gamma \backslash \mathcal{H}_n^* \) is a normal projective variety over \( \mathbb{C} \), which is called the Satake–Baily–Borel or minimal compactification of \( \Gamma \backslash H_n \). (The references given above are not the historical ones mainly for \( \mathcal{H}_n^* \) and \( \Gamma \backslash \mathcal{H}_n^* \), such as [Sat56], [Bai58], [Car58], and [SC58]—see [Sat01, Sec. 2] for a nice review. Rather, they are more recent ones that also work for our later examples.)

Each \( \begin{pmatrix} \infty_r & \cdot \\
 H_{n-r} & \cdot \\
\end{pmatrix} \), or rather \( \left\{ \begin{pmatrix} 1_r & Z \\
 0_r & 1_{n-r} \\
\end{pmatrix} : Z \in \mathcal{H}_{n-r} \right\} \) in generalized projective coordinates (see (4.1.3.13)), is stabilized by a maximal parabolic subgroup of
G(\mathbb{Q}) of the form

\begin{equation}
P(\mathbb{Q}) = \left\{ \begin{pmatrix} * & * & * \\
* & * & * \\
* & * & * 
\end{pmatrix} \right\}
\end{equation}

because

\begin{equation}
\begin{pmatrix}
* & * & * \\
* & * & * \\
* & * & *
\end{pmatrix}
\begin{pmatrix}
1 & * \\
0 & * \\
1 & 1
\end{pmatrix}
= \begin{pmatrix}
* & * \\
0 & 1 \\
1 & 1
\end{pmatrix}
\sim
\begin{pmatrix}
1 & * \\
0 & 1 \\
1 & 1
\end{pmatrix},
\end{equation}

where the sizes of the block matrices are given by \(r, n-r, r, \) and \(n-r,\) both horizontally and vertically. This \(P(\mathbb{Q})\) is the stabilizer of the totally isotropic \(\mathbb{Q}\)-subspace

\begin{equation}
\left\{ \begin{pmatrix}
x \\
0 \\
0
\end{pmatrix} : x \in \mathbb{Q}^r \right\}
\end{equation}

of

\begin{equation}
\mathbb{Q}^{2n} \cong \mathbb{Q}^r \oplus \mathbb{Q}^{n-r} \oplus \mathbb{Q}^r \oplus \mathbb{Q}^{n-r}.
\end{equation}

We have

\begin{equation}
P(\mathbb{Q}) = L(\mathbb{Q}) \ltimes U(\mathbb{Q}),
\end{equation}

where

\begin{equation}
L(\mathbb{Q}) = \left\{ \begin{pmatrix}
X & A & B \\
A & X^{-1} & C \\
B & C & D
\end{pmatrix} : X \in \text{GL}_r(\mathbb{Q}), \left( \begin{array}{cc}
A & B \\
C & D
\end{array} \right) \in \text{Sp}_{2n-2r}(\mathbb{Q}) \right\}
\end{equation}

\[ \cong \text{GL}_r(\mathbb{Q}) \ltimes \text{Sp}_{2n-2r}(\mathbb{Q}) \]

is a Levi subgroup, and where

\begin{equation}
U(\mathbb{Q}) = \left\{ \begin{pmatrix}
1 & * & * \\
1 & * \\
1 & 1
\end{pmatrix} \right\}
\end{equation}

is the unipotent radical. Note that \(L(\mathbb{Q})\) acts on the rational boundary component \(\mathcal{H}_{2n-2r}\) via the second factor \(\text{Sp}_{2n-2r}(\mathbb{Q})\). This is the so-called Hermitian part of the Levi subgroup \(L(\mathbb{Q})\). The unipotent radical \(U(\mathbb{Q})\) is an extension

\begin{equation}1 \to W(\mathbb{Q}) \to U(\mathbb{Q}) \to V(\mathbb{Q}) \to 1,
\end{equation}

where

\begin{equation}
W(\mathbb{Q}) = \left\{ \begin{pmatrix}
1 & Y \\
1 & 1 \\
1 & 1
\end{pmatrix} : Y \in \text{Sym}_r(\mathbb{Q}) \right\} \cong \text{Sym}_r(\mathbb{Q})
\end{equation}
is the center of $U^{(r)}(\mathbb{Q})$, and where

\[(4.1.3.24) \quad V^{(r)}(\mathbb{Q}) \cong \mathbb{Q}^{2n-2r}\]

(by viewing Sym$_r(\mathbb{Q})$ and $\mathbb{Q}^{2n-2r}$ as commutative algebraic groups over $\mathbb{Q}$).

### 4.1.4. Cases with $D = \mathcal{H}_{a,b}$, with $a \geq b$.

Suppose $E$ is some imaginary quadratic extension field of $\mathbb{Q}$. Suppose $G$ is a group scheme over $\mathbb{Q}$ defined by

\[(4.1.4.1) \quad G(R) := \{ g \in \text{GL}_{a+b}(E \otimes \mathbb{Q} R) : \mathcal{G}J_{a,b}g = J_{a,b} \},\]

for each $\mathbb{Q}$-algebra $R$, where $J_{a,b}$ is some skew-Hermitian matrix as in (3.2.2.2) with $S$ a skew-Hermitian matrix with entries in $E$. (We do not need to choose any $S$ when $a = b$.) Then $G(\mathbb{R})$ is the group $U'_{a,b}$ defined by $J_{a,b}$ as in (3.2.2.1), which acts on $\mathcal{H}_{a,b}$ as in Section 3.2.2.

In this case, the rational boundary components of $\mathcal{H}_{a,b}$ (with respect to the choice of $G$) are given by elements of the $G(\mathbb{Q})$-orbits of

\[(4.1.4.2) \quad (\infty_r \mathcal{H}_{a-r,b-r}) \cong \mathcal{H}_{a-r,b-r},\]

which means

\[(4.1.4.3) \quad \left\{ \begin{pmatrix} 1_r & Z \\ W & 0_r \\ 0_r & 1_{b-r} \end{pmatrix} : \begin{pmatrix} Z \\ W \end{pmatrix} \in \mathcal{H}_{a-r,b-r} \right\}\]

in generalized projective coordinates (cf. Remark 3.1.1.4), for $r = 1, \ldots, b$. (When $r = 0$, we just get $\mathcal{H}_{a,b}$ itself.)

With the Satake topology (see the reference given in Section 4.1.3, which we shall not repeat) on the set

\[(4.1.4.4) \quad \mathcal{H}^*_r := \mathcal{H}_{a,b} \cup \bigcup_{1 \leq r \leq b} G(\mathbb{Q}) \cdot \mathcal{H}_{a-r,b-r},\]

where $\mathcal{H}_{a-r,b-r}$ stands for $(\infty_r \mathcal{H}_{a-r,b-r})$ as in (4.1.4.2). $\mathcal{H}_{a-r,b-r}$ lies in the closure of $\mathcal{H}_{a-s,b-s}$ exactly when $r \geq s$, and the quotient $\Gamma \backslash \mathcal{H}^*_r$ is a compactification of $\Gamma \backslash \mathcal{H}_{a,b}$ as a topological space, for every arithmetic subgroup $\Gamma$ of $G(\mathbb{Q})$. Again, according to [BB66], $\Gamma \backslash \mathcal{H}^*_r$ is a normal projective variety over $\mathbb{C}$, which is the minimal compactification of $\Gamma \backslash \mathcal{H}_{a,b}$.

Each $(\infty_r \mathcal{H}_{a-r,b-r})$, or rather $\left\{ \begin{pmatrix} 1_r & Z \\ W & 0_r \\ 0_r & 1_{b-r} \end{pmatrix} : \begin{pmatrix} Z \\ W \end{pmatrix} \in \mathcal{H}_{a-r,b-r} \right\}$ in generalized projective coordinates (see (4.1.4.3)), is stabilized by a maximal parabolic
subgroup of $G(\mathbb{Q})$ of the form

\[(4.1.4.5) \quad P^{(r)}(\mathbb{Q}) = \begin{cases} \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \\ \begin{pmatrix} * \\ * \\ * \\ * \end{pmatrix} \end{cases}, \]

where the sizes of the block matrices are given by $r$, $b - r$, $a - b$, $r$, and $b - r$, both horizontally and vertically. This $P^{(r)}(\mathbb{Q})$ is the stabilizer of the totally isotropic $E$-subspace

\[(4.1.4.6) \quad E^{a+b} \cong E^r \oplus E^{b-r} \oplus E^{a-b} \oplus E^r \oplus E^{b-r}.\]

We have

\[(4.1.4.7) \quad P^{(r)}(\mathbb{Q}) = L^{(r)}(\mathbb{Q}) \rtimes U^{(r)}(\mathbb{Q}), \]

where

\[(4.1.4.8) \quad L^{(r)}(\mathbb{Q}) = \begin{cases} \begin{pmatrix} X & A & E & B \\ F & M & G \end{pmatrix} \\ \begin{pmatrix} C & H & D \end{pmatrix} : X \in \text{GL}_r(E), \begin{pmatrix} A & E & B \\ F & M & G \end{pmatrix} \in G^{(r)}(\mathbb{Q}) \end{cases} \]

\[\cong \text{GL}_r(E) \times G^{(r)}(\mathbb{Q})\]

is a Levi subgroup, where $G^{(r)}$ is the analogue of $G$ defined by the smaller matrix

\[(4.1.4.9) \quad J_{a-r,b-r} = \begin{pmatrix} 1 & * & * & * \\ 1 & * \\ 1 & * \\ * & 1 \end{pmatrix} \quad \text{with the same $S$; and where} \]

\[(4.1.4.10) \quad U^{(r)}(\mathbb{Q}) = \begin{cases} \begin{pmatrix} 1 & * & * & * \\ 1 & * \\ 1 & * \end{pmatrix} \end{cases} \]

is the unipotent radical. Note that $L^{(r)}(\mathbb{Q})$ acts on the rational boundary component $H_{a-r,b-r}$ via the second factor $G^{(r)}(\mathbb{Q})$. This is the Hermitian part of the Levi subgroup $L^{(r)}(\mathbb{Q})$. The unipotent radical $U^{(r)}(\mathbb{Q})$ is an extension

\[(4.1.4.11) \quad 1 \to W^{(r)}(\mathbb{Q}) \to U^{(r)}(\mathbb{Q}) \to V^{(r)}(\mathbb{Q}) \to 1,\]
Res algebraic group over $\mathbb{Q}$, although we need to work with coordinates over $E$ automatically compact, for every arithmetic subgroup $\Gamma$ of $G(\mathbb{Q})$ some integers $a, b, r$ for a skew-Hermitian matrix $J$ extension of a totally real extension $F$ above definition of $G$ with a CM extension (namely, a totally imaginary quadratic $H$ we note that the overall picture is very similar to the case of $\text{Res}_{E/Q} \GL_r(E)$, and with the matrix $S$ when $a > b$.

Let us mention two important variants which frequently appear in the literature.

The first variant replaces the imaginary quadratic extension field $E$ of $\mathbb{Q}$ in the above definition of $G$ with a CM extension (namely, a totally imaginary quadratic extension of a totally real extension) $F$ of $\mathbb{Q}$, and replaces $J$ with a nondegenerate skew-Hermitian matrix $J$ over $\mathcal{O}_F$ such that $-iJ$ defines a Hermitian pairing of signatures $(a_1, b_1), (a_2, b_2), \ldots, (a_d, b_d)$ at the $d$ real places of the maximal totally real subextension $F^+$ of $F$, where $d := [F^+ : \mathbb{Q}]$, for some integers $a_j \geq b_j \geq 0$ and $n \geq 0$ such that $a_j + b_j = n$, for all $j = 1, \ldots, d$. Then

$$G(\mathbb{R}) \cong U'_{a_1, b_1} \times U'_{a_2, b_2} \times \cdots \times U'_{a_d, b_d}$$

acts on $\mathcal{D} \cong \mathcal{H}_{a_1, b_1} \times \mathcal{H}_{a_2, b_2} \times \cdots \times \mathcal{H}_{a_d, b_d}$, and the rational boundary components of $\mathcal{D}$ are given by elements of the $G(\mathbb{Q})$-orbits of some $\mathcal{D}^{(r)} \cong \mathcal{H}_{a - r, b_1 - r} \times \mathcal{H}_{a_2 - r, b_2 - r} \times \cdots \times \mathcal{H}_{a_d - r, b_d - r}.$

for $r = 1, \ldots, \min(b_1, b_2, \ldots, b_d)$. Note that there is no $r$ at all if at least one of $b_1, b_2, \ldots, b_d$ is zero.

In the special case where $(a_1, b_1) = (a, b)$ but $(a_j, b_j) = (n, 0)$ for all $j \neq 1$, for some integers $a \geq b \geq 0$ and $n = a + b \geq 0$, we simply have

$$\mathcal{D} \cong \mathcal{H}_{a, b}.$$ 

If $d = 1$, then this is just as before. If $d \geq 2$, then $\mathcal{D}^* = \mathcal{D}$ and the quotient $\Gamma \backslash \mathcal{D}$ is automatically compact, for every arithmetic subgroup $\Gamma$ of $G(\mathbb{Q})$. Whether $d \geq 2$ or not, when $(a, b) = (2, 1)$ (and when (4.1.4.18) holds), the quotient $\Gamma \backslash \mathcal{D}$ is called a Picard modular surface.

The second variant is a further generalization of the above first variant, by replacing $F$ with a central simple algebra $B$ over a CM extension over $\mathbb{Q}$ that is equipped with a positive involution $\star$ (namely, an anti-automorphism of order 2) such that $\text{tr}_{B/Q}(xx^*) > 0$ for every $x \not= 0$ in $B$, and by defining $G$ using a nondegenerate skew-Hermitian pairing over a finitely generated $B$-module. Then we still have an isomorphism as in (4.1.4.16), and the rational boundary components of $\mathcal{D}$ are still of the same form as in (4.1.4.17), but the values of $r$ allowed in (4.1.4.17) are more restrictive, depending on the possible $\mathbb{Q}$-dimensions of totally isotropic...
If $B$ is a division algebra and the pairing is defined over a simple $B$-module, then no $r > 0$ is allowed, in which case $\Gamma \backslash \mathcal{D}$ is again automatically compact, for every arithmetic subgroup $\Gamma$ of $G(\mathbb{Q})$.

4.1.5. Cases with $\mathcal{D} = \mathcal{H}_{SO_{2n}}^\ast$. Starting with this case, we will be less explicit about the choices of $G$. There exists a group scheme $G$ over $\mathbb{Q}$ such that $G(\mathbb{R}) \cong SO_{2n}^\ast$, such that the rational boundary components of $\mathcal{H}_{SO_{2n}}^\ast$ are given by elements of the $G(\mathbb{Q})$-orbits of some $\mathcal{D}^{(r)} \cong \mathcal{H}_{SO_{2n-4r}}^\ast$ (which we shall abusively denote as $\mathcal{H}_{SO_{2n-4r}}^\ast$, with $r = 1, \ldots, \lfloor \frac{n}{2} \rfloor$, and such that, with the Satake topology on $H_{SO_{2n}}^\ast$:

\begin{equation}
\mathcal{H}_{SO_{2n}}^\ast := \mathcal{H}_{SO_{2n}} \cup \left( \bigcup_{1 \leq r \leq \lfloor \frac{n}{2} \rfloor} G(\mathbb{Q}) \cdot \mathcal{H}_{SO_{2n-4r}}^\ast \right),
\end{equation}

$\mathcal{H}_{SO_{2n-4r}}^\ast$ lies in the closure of $\mathcal{H}_{SO_{2n-4r}}^\ast$ exactly when $r \geq s$, and the quotient $\Gamma \backslash \mathcal{H}_{SO_{2n}}^\ast$ gives the minimal compactification of $\Gamma \backslash \mathcal{H}_{SO_{2n}}^\ast$, for every arithmetic subgroup $\Gamma$ of $G(\mathbb{Q})$.

When $n = 2k$, an example of $G$ is given by

\begin{equation}
G(\mathbb{R}) := \{ g \in GL_{2k}(\mathcal{O} \otimes \mathbb{R}) : t^g J = J \},
\end{equation}

for each $\mathbb{Q}$-algebra $R$, where $B$ is a quaternion algebra over $\mathbb{Q}$ such that $B \otimes \mathbb{R} \cong \mathbb{H}$, and where $J_k = \begin{pmatrix} 1_k & 0 \\ -1_k & 0 \end{pmatrix}$ is as before. Then we can describe the rational boundary components $\mathcal{H}_{SO_{4k-4r}}^\ast$ of $\mathcal{H}_{SO_{4k}}^\ast$ as

\begin{equation}
\left\{ \begin{pmatrix} \infty_r \\ Z \end{pmatrix} : Z \in \text{Herm}_{k-r}(\mathbb{H}) \otimes \mathbb{C}, \text{Im}(Z) > 0 \right\},
\end{equation}

which means

\begin{equation}
\left\{ \begin{pmatrix} 1_r \\ 0_r \\ Z \end{pmatrix} : Z \in \text{Herm}_{k-r}(\mathbb{H}) \otimes \mathbb{C}, \text{Im}(Z) > 0 \right\}
\end{equation}

in generalized projective coordinates (cf. Remark 3.1.1.4), for $r = 1, \ldots, k$. Each $\mathcal{H}_{SO_{4k-4r}}^\ast$ as above is stabilized by a maximal parabolic subgroup of $G(\mathbb{Q})$ of the form

\begin{equation}
P^{(r)}(\mathbb{Q}) = \left\{ \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \right\}
\end{equation}

because

\begin{equation}
\begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} * & * \\ 0 & * \\ 0 & * \end{pmatrix} \sim \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},
\end{equation}

where the sizes of the block matrices are given by $r, n - r, r$, and $n - r$, both horizontally and vertically. This $P^{(r)}(\mathbb{Q})$ is the stabilizer of the totally isotropic
$B$-submodule

\begin{equation}
(4.1.5.7) \quad \left\{ \begin{pmatrix} x \\ 0 \\ 0 \\ 0 \end{pmatrix} : x \in B^r \right\}
\end{equation}

of

\begin{equation}
(4.1.5.8) \quad B^{2n} \cong B^r \oplus B^{n-r} \oplus B^r \oplus B^{n-r},
\end{equation}

with left $B$-module structures given by having each $b \in B$ act by right multiplication of its conjugate $\overline{b}$, so that they do not intervene with left multiplications of matrix entries. We have

\begin{equation}
(4.1.5.9) \quad P^{(r)}(Q) = L^{(r)}(Q) \ltimes U^{(r)}(Q),
\end{equation}

where

\begin{equation}
(4.1.5.10) \quad L^{(r)}(Q) = \left\{ \begin{pmatrix} X & A & B \\ \Gamma^{-1} & C & D \end{pmatrix} : X \in \text{GL}_r(B), \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in G^{(r)}(Q) \right\}
\end{equation}

is a Levi subgroup, where $G^{(r)}$ is the analogue of $G$ defined by the smaller matrix

\begin{equation}
(4.1.5.11) \quad J_{k-r} = \begin{pmatrix} -1_{k-r} & 1_{k-r} \\ \end{pmatrix};
\end{equation}

and where

\begin{equation}
(4.1.5.12) \quad U^{(r)}(Q) = \left\{ \begin{pmatrix} 1 & * & * & * \\ 1 & * & 1 & 1 \\ * & 1 \end{pmatrix} \right\}
\end{equation}

is the unipotent radical. Note that $L^{(r)}(Q)$ acts on the rational boundary component $\mathcal{H}_{SO_{4n-4r}^*}$ via the second factor $G^{(r)}(Q)$. This is the Hermitian part of the Levi subgroup $L^{(r)}(Q)$. The unipotent radical $U^{(r)}(Q)$ is an extension

\begin{equation}
(4.1.5.13) \quad 1 \to W^{(r)}(Q) \to U^{(r)}(Q) \to V^{(r)}(Q) \to 1,
\end{equation}

where

\begin{equation}
(4.1.5.14) \quad W^{(r)}(Q) = \left\{ \begin{pmatrix} 1 & Y \\ 1 & 1 \end{pmatrix} : Y \in \text{Herm}_r(B) \right\} \cong \text{Herm}_r(B)
\end{equation}

is the center of $U^{(r)}(Q)$, and where

\begin{equation}
(4.1.5.15) \quad V^{(r)}(Q) \cong B^{2k-2r}
\end{equation}

(by viewing $\text{Herm}_r(B)$ and $B^{2k-2r}$ as commutative algebraic groups over $Q$). We note that, again, the overall picture is very similar to the one in Section 4.1.3 although we need to work with coordinates over $B$ and hence with $\text{GL}_r(B)$ as an algebraic group over $Q$. 
4.1.6. Cases with $D^+ = \mathcal{H}^+_{SO_{a,2}}$, with $a \geq 2$. Suppose $G$ is a group scheme over $\mathbb{Q}$ defined by some $\mathbb{Q}$-valued symmetric bilinear form $Q$ on $\mathbb{Q}^{a+2}$ which has signature $(a,2)$ over $\mathbb{R}$, so that $G(\mathbb{R}) \cong SO_{a,2}(\mathbb{R})$ and $G(\mathbb{R})^+ \cong SO_{a,2}(\mathbb{R})^+$. (The precise choice of $Q$ will not be important for our exposition.)

The maximal parabolic subgroups of $G(\mathbb{Q}) = SO(\mathbb{Q}^{a+2}, \mathbb{Q})$, where $SO(\mathbb{Q}^{a+2}, \mathbb{Q})$ denotes the subgroup of $SL_{a+2}(\mathbb{Q})$ consisting of elements preserving the pairing $Q$ (and we shall adopt similar notation in what follows), are in bijection with the nonzero subspaces $I$ of $\mathbb{Q}^{a+2}$ that are totally isotropic with respect to the bilinear form $Q$. Then we must have $1 \leq \dim_{\mathbb{Q}}(I) \leq 2$ because the signature of $Q$ is $(a,2)$ over $\mathbb{R}$. (But it might happen that no such $I$ exists.) For each such $I$, we have a parabolic subgroup of $G(\mathbb{Q})$ of the form

\begin{equation}
P_I(\mathbb{Q}) \cong L_I(\mathbb{Q}) \rtimes U_I(\mathbb{Q}),
\end{equation}

where

\begin{equation}
L_I(\mathbb{Q}) \cong SO(I^\perp/I, Q|_{I^\perp/I}) \times GL_Q(I)
\end{equation}
is the Levi quotient (which is also noncanonically a subgroup), and where $U_I(\mathbb{Q})$ is the unipotent radical which fits into a short exact sequence

\begin{equation}
1 \to \wedge^2 I \to U_I(\mathbb{Q}) \to \text{Hom}_{\mathbb{Q}}(I^\perp/I, I) \to 1.
\end{equation}

When $\dim_{\mathbb{Q}}(I) = 1$, we have

\begin{equation}
L_I(\mathbb{R}) \cong SO_{a-1,1}(\mathbb{R}) \times GL_1(\mathbb{R}),
\end{equation}

and so $L_I(\mathbb{R}) \cap G(\mathbb{R})^+$ acts trivially on some zero-dimensional rational boundary component

\begin{equation}
\mathcal{H}_I^+ \cong \mathcal{H}_0.
\end{equation}

When $\dim_{\mathbb{Q}}(I) = 2$, we have

\begin{equation}
L_I(\mathbb{R}) \cong SO_{a-2}(\mathbb{R}) \times GL_2(\mathbb{R}),
\end{equation}

and so $L_I(\mathbb{R}) \cap G(\mathbb{R})^+$ acts via the usual Möbius transformation of

\begin{equation}
\text{GL}_2(\mathbb{R})^+ = \{ g \in \text{GL}_2(\mathbb{R}) : \det(g) > 0 \}
\end{equation}
on some one-dimensional rational boundary component

\begin{equation}
\mathcal{H}_I^+ \cong \mathcal{H}_1.
\end{equation}

For example, suppose (for simplicity) that $Q$ is defined by the symmetric matrix

\begin{equation}
\begin{pmatrix}
1_{a-2} & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 1 \\
0 & 1 & 0 \\
\end{pmatrix},
\end{equation}

where the sizes of the block matrices are given by $a - 2$, 1, 1, 1, and 1. If

\begin{equation}
I = \begin{pmatrix}
0 \\
0 \\
* \\
0 \\
\end{pmatrix}
\end{equation}
which is one-dimensional, then

\begin{equation}
\mathcal{P}_I(\mathbb{Q}) = \left\{ \begin{pmatrix}
* & * & * & * \\
* & * & * & * \\
* & * & * & * \\
* & * & * & *
\end{pmatrix} \right\} \cong \mathcal{L}_I(\mathbb{Q}) \ltimes \mathcal{U}_I(\mathbb{Q}),
\end{equation}

where

\begin{equation}
\mathcal{L}_I(\mathbb{Q}) = \left\{ \begin{pmatrix}
* & * & * & * \\
* & * & * & * \\
* & * & * & * \\
X & X^{-1}
\end{pmatrix} : \begin{pmatrix}
* & * & * & * \\
* & * & * & *
\end{pmatrix} \in \text{SO}_{a-1,1}(\mathbb{Q}),
X \in \mathbb{Q}^\times \right\}
\end{equation}

and

\begin{equation}
\mathcal{U}_I(\mathbb{Q}) = \left\{ \begin{pmatrix}
1 & 1 & x \\
1 & -y & z \\
x & -z & 1 \\
y & y & 1
\end{pmatrix} : x \in \mathbb{Q}^{a-2},
y, z \in \mathbb{Q} \right\};
\end{equation}

and \(\mathcal{P}_I(\mathbb{Q})\) stabilizes

\begin{equation}
\mathcal{H}_I^+ = \left\{ \begin{pmatrix} 0 \\
0 \\
0 \\
1 \end{pmatrix} \right\}
\end{equation}

because

\begin{equation}
\begin{pmatrix}
* & * & * & * \\
* & * & * & * \\
* & * & * & * \\
* & * & * & *
\end{pmatrix} \begin{pmatrix} 0 \\
0 \\
0 \\
1 \end{pmatrix} = \begin{pmatrix} 0 \\
0 \\
0 \\
1 \end{pmatrix} \sim \begin{pmatrix} 0 \\
0 \\
0 \\
0 \end{pmatrix}
\end{equation}

in projective coordinates. Alternatively, if

\begin{equation}
I = \left\{ \begin{pmatrix} 0 \\
* \\
0 \\
0 \end{pmatrix} \right\},
\end{equation}

which is two-dimensional, then

\begin{equation}
\mathcal{P}_I(\mathbb{Q}) = \left\{ \begin{pmatrix}
* & * & * & * \\
* & * & * & * \\
* & * & * & * \\
* & * & * & *
\end{pmatrix} \right\} \cong \mathcal{L}_I(\mathbb{Q}) \ltimes \mathcal{U}_I(\mathbb{Q}),
\end{equation}
where

\[
L_I(Q) = \left\{ \begin{pmatrix} X \\ a & b \\ \alpha & d \end{pmatrix} : \begin{pmatrix} a & b \\ \alpha & d \end{pmatrix}^{-1} = \begin{pmatrix} a & b \\ \alpha & d \end{pmatrix} \in GL_2(Q) \right\}
\]

and

\[
U_I(Q) = \left\{ \begin{pmatrix} 1 & 0 \\ -x & 1 \\ 0 & 1 \\ y \\ \frac{1}{2} \end{pmatrix} : w \in Q, x, y \in Q^{n-2} \right\};
\]

and \(P_I(Q) \cap G(R)^+\) stabilizes

\[
H^+_I = \left\{ \begin{pmatrix} z \\ 0 \\ 1 \end{pmatrix} : z \in \mathbb{C}, \text{Im} z > 0 \right\}
\]

because

\[
\begin{pmatrix} \ast & \ast & \ast & \ast \\ \ast & a & b & \ast \\ \ast & \ast & \ast & \ast \\ \ast & c & d & \ast \end{pmatrix} \begin{pmatrix} 0 \\ z \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ az + b \\ 0 \\ cz + d \end{pmatrix} \sim \begin{pmatrix} 0 \\ (az + b)(cz + d)^{-1} \\ 0 \\ 1 \end{pmatrix}
\]

in projective coordinates.

There is a systematic way to combine these into a union

\[
H^+_I \cup \bigcup_{0 \neq I \text{ totally isotropic}} H^+_I
\]

equipped with Satake topology, such that \(H^+_I\) lies in the closure of \(H^+_I\) (for nonzero \(I\) and \(I'\)) exactly when \(I \subset I'\) (which is qualitatively quite different from the previous cases in Sections 3.1.3, 3.1.4, and 3.1.5 where rational boundary components associated with larger totally isotropic subspaces are smaller) and such that the quotient \(\Gamma / H^+_I\) gives the minimal compactification of \(\Gamma / H^+_I\) for every arithmetic subgroup \(\Gamma\) of \(G(Q)\) that is contained in \(G(R)^+\).

4.1.7. Cases with \(D = H_{E_6}\) or \(D = H_{E_7}\). We shall be very brief in these two cases. For \(H_{E_6}\), there exists a group scheme \(G\) over \(Q\) such that \(G(R)\) is as in Section 3.6.2 and such that the rational boundary components of \(H_{E_6}\) are isomorphic to either \(H_{5,1}\) or \(H_6\). For \(H_{E_7}\), there exists a group scheme \(G\) over \(Q\) such that \(G(R)\) is as in Section 3.5.3 and such that the rational boundary components of \(H_{E_7}\) are isomorphic to either \(H_{SO_{10,2}}^+, H_1\), or \(H_0\). In each of these two cases, by [PR06, Sec. 3, Prop. 3], there exists some \(G\) such that \(G(R)\) is as above and such that \(\text{rank}_Q G_Q = \text{rank}_R G_R\); or, in other words, \(G_Q\) is as split as \(G_R\) allows. In both
cases, we have the minimal compactifications, as usual, for arithmetic subgroups $\Gamma$ of $G(\mathbb{Q})$.

4.2. Compactifications and algebraicity.

4.2.1. Overview. In all cases considered so far, we have the following diagram of useful compactifications of $\Gamma \backslash \mathcal{D}$, with canonical morphisms between them denoted by solid arrows:

(4.2.1.1) $\xymatrix{ (\Gamma \backslash \mathcal{D})^\text{tor} \ar[r] & (\Gamma \backslash \mathcal{D})^\text{BS} \ar[d] \ar[r] & (\Gamma \backslash \mathcal{D})^\text{RBS} \ar[d] \ar[r] & (\Gamma \backslash \mathcal{D})^\text{min} \ar[l] \ar[d] \ar[r] & (\Gamma \backslash \mathcal{D})^\text{BS} \ar[d] \ar[r] & (\Gamma \backslash \mathcal{D})^\text{RBS} \ar[d] \ar[r] & (\Gamma \backslash \mathcal{D})^\text{min} \ar[l] }$

Let us explain the objects in this diagram.

4.2.2. $(\Gamma \backslash \mathcal{D})^\text{min} = \Gamma \backslash \mathcal{D}^*$. This is the Satake–Baily–Borel or minimal compactification of $\Gamma \backslash \mathcal{D}$, which is a normal projective variety over $\mathbb{C}$ (see [BB66]). Here $\mathcal{D}^*$ is the union of $\mathcal{D}$ with its rational boundary components, equipped with the Satake topology, as explained in Section 4.1. The transition from the analytic quotient $\Gamma \backslash \mathcal{D}^*$ to an algebraic variety is achieved by first endowing $\Gamma \backslash \mathcal{D}^*$ with the structure of a complex analytic space (see [Ser56, §1]) in which the closed complement of the open subset $\Gamma \backslash \mathcal{D}$ of $\Gamma \backslash \mathcal{D}^*$ is a closed complex analytic subspace; and then by constructing a closed immersion of $\Gamma \backslash \mathcal{D}^*$ into a projective space $\mathbb{P}^N(\mathbb{C})$ for some $N$, and by applying Chow’s theorem (or “GAGA” for closed complex analytic subspaces of $\mathbb{P}^N(\mathbb{C})$, showing that they are all complex analytifications of algebraic subvarieties—see [Ser56, §3, 19, Prop. 13]). Then the open subspace $\Gamma \backslash \mathcal{D}$ of $(\Gamma \backslash \mathcal{D})^\text{min} = \Gamma \backslash \mathcal{D}^*$ is a quasi-projective variety over $\mathbb{C}$, by applying Chow’s theorem again to the closed complement. (This is the only known general method to show that $\Gamma \backslash \mathcal{D}$ is an algebraic variety.) Moreover, it is proved in [Bor72] that any holomorphic map from a quasi-projective variety to $\Gamma \backslash \mathcal{D}$ is the complex analytification of a morphism of algebraic varieties. In particular, the structure of $\Gamma \backslash \mathcal{D}$ as an algebraic variety is unique up to isomorphism. (These are the key results used in the proof of Theorem 2.4.1.)

The compactification $(\Gamma \backslash \mathcal{D})^\text{min}$ is canonical and does not depend on any choices, but is highly singular in general. As we have seen in the examples in Section 4.1 the boundary $(\Gamma \backslash \mathcal{D})^\text{min} - (\Gamma \backslash \mathcal{D})$ has a high codimension in $(\Gamma \backslash \mathcal{D})^\text{min}$ in general. When the connected components of $\mathcal{D}$ are irreducible in the sense explained in the beginning of Section 3 but are not one-dimensional, and when $\Gamma \backslash \mathcal{D}$ is not already compact, this codimension is always at least two. The intuition is that we are collapsing a lot into a small neighborhood of each boundary point.

4.2.3. $(\Gamma \backslash \mathcal{D})^\text{tor}$. This is the toroidal compactification of $\Gamma \backslash \mathcal{D}$ (see [AMRT75] or [AMRT10], and see also [Mum75]), which is an algebraic space depending on the choice of some combinatorial data $\Sigma$ called a cone decomposition (or fan). (More precisely, $\Sigma$ is a compatible collection of cone decompositions, but we prefer not
to go into the details here.) This means \((\Gamma\backslash D)^{\text{tor}}\) is not a canonical compactification of \(\Gamma\backslash D\) in the sense that it does not just depend on \(\Gamma\backslash D\). Nevertheless, it depends only on \(\Gamma\backslash D\) and \(\Sigma\), and there exist choices of \(\Sigma\) such that \((\Gamma\backslash D)^{\text{tor}}\) is a smooth projective variety, and such that the boundary \((\Gamma\backslash D)^{\text{tor}} - (\Gamma\backslash D)\) is a simple normal crossings divisor. Such choices can always be achieved by replacing any given \(\Sigma\) with a refinement. In particular, we obtain meaningful smooth projective compactifications by paying the price of working with the noncanonical \(\Sigma\)'s. (As a comparison, we have no precise control on the smooth compactifications obtained by resolving the singularities of \((\Gamma\backslash D)^{\text{min}}\) using Hironaka’s general theory in [Hir64a, Hir64b].)

As in the case of \((\Gamma\backslash D)^{\text{min}}\), the boundary of \((\Gamma\backslash D)^{\text{tor}}\) is stratified by locally closed subsets corresponding to \(\Gamma\)-orbits of the proper parabolic subgroups \(P\) of the algebraic group \(G\) over \(\mathbb{Q}\) whose pullback to every \(\mathbb{Q}\)-simple (almost) factor \(G'\) of \(G\) is either the whole \(G'\) or a maximal parabolic subgroup \(P'\). This stratification is further refined by a stratification in terms of the cones in the cone decomposition \(\Sigma\). The incidence relations of the strata in this finer stratification is dual to the incidence relations of the cones in the cone decomposition \(\Sigma\). The canonical map \((\Gamma\backslash D)^{\text{tor}}_\Sigma \to (\Gamma\backslash D)^{\text{min}}\) in (4.2.1.1) respects such stratifications.

One subtlety is that, except in very special cases, no choice of \(\Sigma\) can be compatible with all Hecke correspondences. For arithmetic applications, it is necessary to systematically work with refinements of \(\Sigma\).

4.2.4. \((\Gamma\backslash D)^{\text{BS}}\). This is the Borel–Serre compactification of \(\Gamma\backslash D\) (see [BS73] and [BJ06 Sec. III.5]; see also [Gor05 Sec. 4]), which is a (real analytic) manifold with corners. It can be defined more generally when \(D\) is a Riemannian symmetric space that is not Hermitian, but is almost never an algebraic variety. Already when \(G = \text{SL}_2\) and \(D = \mathbb{H}\) (and when \(\Gamma\) is neat), the compactification \((\Gamma\backslash D)^{\text{BS}}\) is achieved by adding circles to \(D\) (instead of adding points to \(D\) as in the case of \(D^*\)). In general, its boundary is stratified by locally closed subsets corresponding to the \(\Gamma\)-orbits of all proper rational parabolic subgroups \(P\) of the algebraic group \(G\) over \(\mathbb{Q}\), and the construction of the subset associated with \(P\) uses the whole \(P\), instead of just its Levi quotient as in the case of \((\Gamma\backslash D)^{\text{min}}\).

The most important feature of \((\Gamma\backslash D)^{\text{BS}}\) is that the canonical open embedding \(\Gamma\backslash D \hookrightarrow (\Gamma\backslash D)^{\text{BS}}\) is a homotopy equivalence, which makes \((\Gamma\backslash D)^{\text{BS}}\) useful for studying the cohomology of \(\Gamma\backslash D\), and also that of the arithmetic subgroup \(\Gamma\) of \(G(\mathbb{Q})\), which are closely related to the theory of automorphic representations and Eisenstein series (see, for example, [BW00], [Sch90a], [Fra98], [FS98], and [LS04]). But let us remark here, as an addendum to Section 4.2.2 above, that cuspidal automorphic representations are more closely related to the \(L^2\)-cohomology of \(\Gamma\backslash D\) and this \(L^2\)-cohomology is canonical isomorphic to the intersection cohomology of \((\Gamma\backslash D)^{\text{min}}\). This was conjectured by Zucker (see [Zuc82 (6.20)]) and proved by Looijenga (see [Loo88]), Saper–Stern (see [SS90]), and Looijenga–Rapoport (see [LR91]).

4.2.5. \((\Gamma\backslash D)^{\text{RBS}}\). This is the reductive Borel–Serre compactification of \(\Gamma\backslash D\) (see [Zuc82 Sec. 4], [GHM94 Part I], and [BJ06 Sec. III.6]; see also [Gor05 Sec. 5]), which is similar to \((\Gamma\backslash D)^{\text{BS}}\) in the following aspects: It can also be defined more generally when \(D\) is a Riemannian symmetric space that is not Hermitian, and its boundary is also stratified by locally closed subsets corresponding to the \(\Gamma\)-orbits of all proper rational parabolic subgroups \(P\) of \(G\). But there is a key difference—the
subset associated with \( P \) is constructed using only the (reductive) Levi quotient \( L \) of \( P \). The reductive Borel-Serre compactification \((\Gamma \backslash \mathcal{D})^{RBS}\) is useful for studying, for example, the weighted cohomology of \( \Gamma \) (see [GHM94] Parts II-IV; see also [Gor05] Sec. 6.8).

4.2.6. Models over number fields or their rings of integers. Although \((\Gamma \backslash \mathcal{D})^{RBS}\) and \((\Gamma \backslash \mathcal{D})^{RHS}\) are very useful analytic objects, when it comes to applications requiring algebraic varieties (or at least algebraic spaces or algebraic stacks) defined over \( \mathbb{Q} \), we need to work with \((\Gamma \backslash \mathcal{D})^{\min}\) and \((\Gamma \backslash \mathcal{D})^{\tor}_{\Sigma}\), the last being perhaps the most complicated. It is known that \((\Gamma \backslash \mathcal{D})^{\min}\) and \((\Gamma \backslash \mathcal{D})^{\tor}_{\Sigma}\) also have models over the same number fields over which \((\Gamma \backslash \mathcal{D})^{\min}\) is defined (see [Pin89]), and they extend to good models over the integers in many important special cases (see Sections 5.1.4 and 5.2.4 below).

All assertions so far in this Section 4.2 have their counterparts for Shimura varieties; i.e., the double coset spaces \((\Gamma \backslash \mathcal{D})\) does not extend to a vector bundle over \((\Gamma \backslash \mathcal{D})\). Nevertheless, the automorphic vector bundle \( W \) extends to a good model over the integers in many important special cases (see \[Gor05, Sec. 6.8\]).

4.2.7. Algebro-geometric definition of modular forms. The projective coordinates of the minimal compactification \((\Gamma \backslash \mathcal{D})^{\min} = \Gamma \backslash \mathcal{D}^{*}\) are given by (holomorphic) modular forms, namely, holomorphic functions \( f : \mathcal{D} \to \mathbb{C} \) satisfying the following two conditions:

1. Automorphy condition: \( f(\gamma Z) = j(\gamma, Z)f(Z) \), for all \( \gamma \in \Gamma \) and \( Z \in \mathcal{D} \), for some automorphy factor \( j : \Gamma \times \mathcal{D} \to \mathbb{C}^{*} \) defining the weight of \( f \), satisfying the cocycle condition \( j(\gamma' \gamma, Z) = j(\gamma', Z)j(\gamma, Z) \) for all \( \gamma, \gamma' \in \Gamma \) and \( Z \in \mathcal{D} \). (This means \( f \) represents a section of an automorphic line bundle over \( \Gamma \backslash \mathcal{D} \) defined by \( j \). The line bundles used in [BB66] are powers of the canonical bundle, namely, the top exterior power of the cotangent bundle, of the complex manifold \( \Gamma \backslash \mathcal{D} \).)

2. Growth condition: \( f(Z) \) stays bounded near \( \mathcal{D}^{*} - \mathcal{D} \). (When \( \Gamma \) is neat, this implies that the section represented by \( f \) extends to a section of a line bundle over \((\Gamma \backslash \mathcal{D})^{\min} = \Gamma \backslash \mathcal{D}^{*}\).) If \( f \) satisfies the additional condition that \( f(Z) \to 0 \) as \( Z \to \mathcal{D}^{*} - \mathcal{D} \), then \( f \) is called a cusp form.

Ratios of such modular forms \( f \) define modular functions which are in general highly transcendental as functions of \( \mathcal{D} \). Therefore, to give models of \( \Gamma \backslash \mathcal{D} \) and \((\Gamma \backslash \mathcal{D})^{\min}\) over number fields or over their rings of integers, we cannot just naively use the coordinates of \( \mathcal{D} \).

Not all modular forms are defined as above using only automorphy factors valued in \( \mathbb{C}^{*} \). It is possible to consider automorphy factors valued in \( \text{Cent}_{h_{\rho}}(G(\mathbb{C})) \), so that it makes sense to consider modular forms with weights given by all finite-dimensional representations \( W \) of \( \text{Cent}_{h_{\rho}}(G(\mathbb{C})) \). But there is a subtlety here: For such a \( W \), in general, the associated automorphic vector bundle \( W \) over \( \Gamma \backslash \mathcal{D} \) does not extend to a vector bundle over \((\Gamma \backslash \mathcal{D})^{\min}\). (This is consistent with the fact that, in general, the boundary \((\Gamma \backslash \mathcal{D})^{\min} - (\Gamma \backslash \mathcal{D})^{\min}\) has a high codimension in \((\Gamma \backslash \mathcal{D})^{\min}\).) Nevertheless, the automorphic vector bundle \( W \) has canonical extensions (see [Mum77] and [Har89]) to vector bundles \( W^{\text{can}} \) over the toroidal compactifications \((\Gamma \backslash \mathcal{D})^{\tor}_{\Sigma} \), whose sections are independent of \( \Sigma \). (When \( W \) is the
restriction of a representation of \( G(\mathbb{C}) \), the vector bundle \( W \) is equipped with an integrable connection, and its canonical extension \( W^{\text{can}} \) is equipped with an integrable connection with log poles along the boundary and coincides with the one given in \([Del70]\). The global sections or more generally the cohomology classes of such canonical extensions \( W^{\text{can}} \) over \((\Gamma \backslash \mathcal{D})^{\text{tor}}\) then provide an algebro-geometric definition of modular forms with coefficients in \( W \). There are also the subcanonical extensions \( W^{\text{sub}} := W^{\text{can}}(-\infty) \), where \( \infty \) abusively denotes the boundary divisor \((\Gamma \backslash \mathcal{D})^{\text{tor}} - (\Gamma \backslash \mathcal{D})\) with its reduced structure, whose global sections are (roughly speaking) useful for studying cusp forms. The cohomology of \( W^{\text{can}} \) and \( W^{\text{sub}} \) is often called the coherent cohomology of \((\Gamma \backslash \mathcal{D})^{\text{tor}}\). It is this approach that will allow us to define modular forms in mixed or positive characteristics and study congruences among them using integral models. Even in characteristic zero, there have been some results which have only been proved with such an algebro-geometric definition. (See \([Har90]\), \([Mil90]\), \([Lan12b]\), and \([Lan19]\) for some surveys on this topic.)

4.2.8. Mixed Shimura varieties. Just as Shimura varieties parameterize certain variations of Hodge structures with additional structures, there is also a notion of mixed Shimura varieties parameterizing certain variations of mixed Hodge structures with additional structures. Then there is also a theory of canonical models for such mixed Shimura varieties, and even for their toroidal compactifications—see \([Mil90]\) and \([Pin89]\). Many mixed Shimura varieties naturally appear along the boundary of the toroidal compactifications of pure Shimura varieties (i.e., the usual ones associated with reductive groups as in Sections 2.3 and 2.4).

The theory of mixed Shimura varieties is even heavier in notation than the theory of (pure) Shimura varieties, and is beyond the scope of this introductory article. But let us at least mention the following prototypical example. Consider the Shimura varieties \( X_{\mathcal{U}} \) associated to \((G, D) = (\text{GL}_2, \mathcal{H}^\pm)\), where \( \mathcal{H}^\pm := \mathcal{H}^\pm \), which are modular curves. For each integer \( m \geq 1 \), consider the (non-reductive) semidirect product \( \tilde{G} = \text{GL}_2 \ltimes (\text{G}_2^\pm)^m \), with \( G = \text{GL}_2 \) acting diagonally on \((\text{G}_2^\pm)^m\), where \( \text{G}_a \) is the additive group scheme over \( \mathbb{Z} \), so that \( \text{GL}_2(R) \) acts by left multiplication on \((\text{G}_2^\pm)(R) = R^2 \) for every ring \( R \). Consider the corresponding action of \( \tilde{G}(\mathbb{R}) \) on \( \tilde{D} = \mathcal{H}^\pm \times (\mathbb{R}^2)^m \). Then we obtain by considering double coset spaces as in Section 2.2 the mixed Shimura varieties \( \tilde{X}_{\mathcal{U}} \) over \( X_{\mathcal{U}} \) for neat open compact subgroups \( \mathcal{U} \) with image in \( \mathcal{U} \) under the canonical homomorphism \( \tilde{G}(\mathbb{A}^\infty) \rightarrow \text{G}(\mathbb{A}^\infty) \), which are torsors under abelian schemes over \( X_{\mathcal{U}} \) that are isogenous to the \( m \)-fold self-fiber product of the universal elliptic curve over \( X_{\mathcal{U}} \). Specifically, there exists some \( \mathcal{U} \) such that \( \tilde{X}_{\mathcal{U}} \) is exactly isomorphic to the \( m \)-fold self-fiber product of the universal elliptic curve over \( X_{\mathcal{U}} \). The projective smooth toroidal compactifications of such mixed Shimura varieties then provide generalizations of the classical Kuga–Sato varieties, whose cohomology with trivial coefficients are useful for studying the cohomology with nontrivial coefficients over the usual modular curves \( X_{\mathcal{U}} \), or even for the construction of related motives—see, for example, \([Del71a\ Sec. 5]\) and \([Sch90a]\). (We can also reduce questions about the cohomology of Shimura varieties with nontrivial coefficients to questions about the cohomology of mixed Shimura varieties with trivial coefficients in many more general cases—see the end of Section 5.1.4 below.)
5. Integral models

5.1. PEL-type cases. All currently known constructions of integral models of Shimura varieties (which we can describe in some reasonable detail) rely on the important coincidence that, as explained in Section 3.1.5, when \( G = \text{Sp}_{2n} \) and \( \Gamma \) is a congruence subgroup of \( G(\mathbb{Z}) \), the Siegel modular variety \( \Gamma \backslash \mathcal{H}_n \) parameterizes polarized abelian varieties with level structures; and on the fact that we can define moduli problems over the integers that naturally extend such complex varieties. More generally, we can define the so-called PEL moduli problems, which parameterize abelian schemes with additional PEL structures. As we shall see below, the abbreviation PEL stands for polarizations, endomorphism structures, and level structures. This is the main topic of this subsection.

5.1.1. PEL datum. The definition of PEL moduli problems require some linear algebraic data. Suppose we are given an integral PEL datum

\[(O, \star, L, \langle \cdot, \cdot \rangle, h_0),\]

whose entries can be explained as follows:

1. \( O \) is an order in a finite-dimensional semisimple algebra over \( \mathbb{Q} \).
2. \( \star \) is a positive involution of \( O \); namely, an anti-automorphism of order 2 such that \( \text{tr}_{O \otimes \mathbb{R}/\mathbb{R}}(xx^*) > 0 \) for every \( x \neq 0 \) in \( O \otimes \mathbb{R} \).
3. \( L \) is an \( O \)-lattice; namely, a finitely generated free \( \mathbb{Z} \)-module \( L \) with the structure of an \( O \)-module.
4. \( \langle \cdot, \cdot \rangle : L \times L \to \mathbb{Z}(1) \) is an alternating pairing satisfying the condition that

\[\langle bx, y \rangle = \langle x, b^* y \rangle\]

for all \( x, y \in L \) and \( b \in O \). Here

\[\mathbb{Z}(1) := \ker(\exp : \mathbb{C} \to \mathbb{C}^\times) = 2\pi i \mathbb{Z}\]

is the formal Tate twist of \( \mathbb{Z} \).
5. \( h_0 \) is an \( \mathbb{R} \)-algebra homomorphism

\[h_0 : \mathbb{C} \to \text{End}_{O \otimes \mathbb{R}}(L \otimes \mathbb{R}),\]

satisfying the following conditions:

a. For any \( z \in \mathbb{C} \) and \( x, y \in L \otimes \mathbb{R} \), we have \( \langle h_0(z)x, y \rangle = \langle x, h_0(\overline{z})y \rangle \), where \( z \mapsto \overline{z} \) denotes the complex conjugation.

b. The (symmetric) \( \mathbb{R} \)-bilinear pairing \( (2\pi i)^{-1}\langle \cdot, h_0(i)(\cdot) \rangle \) on \( L \otimes \mathbb{R} \) is positive definite.

What do these mean? Given any integral PEL datum as above, we can write down a real torus

\[A_0 := (L \otimes \mathbb{R})/L\]

with a complex structure given by \( h_0 : \mathbb{C} \to \text{End}_{O \otimes \mathbb{R}}(L \otimes \mathbb{R}) \), which can be shown to be an abelian variety (not just a complex torus) with a polarization

\[\lambda_0 : A_0 \to A_0^\vee \cong (L \otimes \mathbb{R})/L^\#.\]
where $A^\vee_0$ is the *dual abelian variety* of $A_0$, with
\begin{equation}
L^\# := \{ x \in L \otimes \mathbb{Q} : (x, y) \in \mathbb{Z}(1), \text{ for all } y \in L \}
\end{equation}
the *dual lattice* of $L$ with respect to $\langle \cdot, \cdot \rangle$, and where the $\lambda_0$ is induced by the natural inclusion $L \subset L^\#$ (see [Mum70, Ch. I, Sec. 2, and Ch. II, Sec. 9] and [Lan12a, Sec. 2.2]; cf. [Lan13, Sec. 1.3.2]). The assumptions that $L$ is an $\mathcal{O}$-lattice and that the compatibility (5.1.1.2) holds imply that we have an *endomorphism structure*
\begin{equation}
i_0 : \mathcal{O} \to \text{End}_C(A_0)
\end{equation}
satisfying the *Rosati condition*; namely, for each $b \in \mathcal{O}$, the diagram
\begin{equation}
\begin{array}{ccc}
A_0 & \xrightarrow{\lambda_0} & A^\vee_0 \\
\downarrow_{i(b)} & & \downarrow_{(i(b))^\vee} \\
A_0 & \xrightarrow{\lambda_0} & A^\vee_0
\end{array}
\end{equation}
is commutative (cf. [Lan13, Sec. 1.3.3]). Moreover, for each integer $n \geq 1$, we have a *principal level-$n$ structure*
\begin{equation}
(\alpha_{0,n} : L/nL \xrightarrow{\sim} A_0[n], \nu_{0,n} : (\mathbb{Z}/n\mathbb{Z})(1) \xrightarrow{\sim} \mu_n),
\end{equation}
where $A_0[n]$ denotes the $n$-torsion subgroup of $A_0$, which is canonically isomorphic to $(\frac{1}{n}L)/L$ because of (5.1.1.5), where $\alpha_{0,n}$ is induced by this canonical isomorphism, and where $\nu_{0,n}$ is induced by the exponential function $\exp : \mathbb{C} \to \mathbb{C}^\times$ (with kernel $\mathbb{Z}(1)$ by definition), so that the diagram
\begin{equation}
\begin{array}{ccc}
(L/nL) \times (L/nL) & \xrightarrow{\langle \cdot, \cdot \rangle} & (\mathbb{Z}/n\mathbb{Z})(1) \\
\downarrow_{\alpha_{0,n} \times \alpha_{0,n}} & & \downarrow_{\nu_{0,n}} \\
A_0[n] \times A_0[n] & \xrightarrow{\lambda_0, \text{Weil pairing}} & \mu_n
\end{array}
\end{equation}
is commutative (see [Lan12a, Sec. 2.2 and 2.3]; cf. [Lan13, Sec. 1.3.6]), and so that such a level structure is *liftable* to all higher levels $n'$ divisible by $n$. (The second entry $\nu_{0,n}$ in $(\alpha_{0,n}, \nu_{0,n})$ is often omitted in the notation, although $\nu_{0,n}$ is not determined by $\alpha_{0,n}$ when $\langle \cdot, \cdot \rangle$ is not a perfect pairing modulo $n$.) Furthermore, the homomorphism $h_0 : \mathbb{C} \to \text{End}_{\mathbb{O} \otimes \mathbb{R}}(L \otimes \mathbb{R})$ defines a *Hodge decomposition*
\begin{equation}
H_1(A_0, \mathbb{C}) \cong L \otimes \mathbb{C} = V_0 \oplus \overline{V}_0,
\end{equation}
where
\begin{equation}
V_0 := \{ x \in L \otimes \mathbb{C} : h_0(z)x = (1 \otimes z)x, \text{ for all } z \in \mathbb{C} \}
\end{equation}
and
\begin{equation}
\overline{V}_0 := \{ x \in L \otimes \mathbb{C} : h_0(z)x = (1 \otimes z)x, \text{ for all } z \in \mathbb{C} \}.
\end{equation}
Then we also have
\begin{equation}
\text{Lie } A_0 \cong V_0
\end{equation}
as $O \otimes \mathbb{C}$-modules. The reflex field $F_0$ for $h_0$ (which is defined in this context without referring to any Shimura datum) is the field of definition of the isomorphism class of $V_0$ as an $O \otimes \mathbb{C}$-module, or more precisely the subfield

$$F_0 := \mathbb{Q}(\text{tr}(b|V_0) : b \in O)$$

of $\mathbb{C}$ (see [Lan13, Sec. 1.2.5]). This is the smallest subfield of $\mathbb{C}$ over which we can formulate a trace or determinantal condition on the Lie algebra of an abelian variety, to ensure that the pullback to $\mathbb{C}$ of its Lie algebra is isomorphic to $V_0$ as an $O \otimes \mathbb{C}$-module (see [Lan13 Sec. 1.2.5 and 1.3.4]).

The upshot is that, by combining all of these, we obtain a tuple

$$(5.1.1.17) \quad (A_0, \lambda_0, i_0, (\alpha_0, n, \nu_0, n)),$$

which is an abelian variety with a polarization, an endomorphism structure, and a level structure, such that Lie $A_0$ satisfies the Lie algebra condition defined by $h_0$. These additional structures form the so-called PEL structures. Thus, writing down an integral PEL datum means writing down such a tuple.

5.1.2. Smooth PEL moduli problems. Given any integral PEL datum as in (5.1.1.1), we can define the associated PEL moduli problems (at varying levels) over (the category of schemes over) $F_0$, or over suitable localizations of the ring of integers $O_{F_0}$. Roughly speaking, they are the smallest moduli problems (over the respective base rings) parameterizing tuples of abelian schemes with additional PEL structures as above, including the one we wrote down in (5.1.1.17). See [Lan13, Sec. 1.4.1] for more details on defining smooth PEL moduli problems over localizations of $O_{F_0}$ of good residue characteristics. (A prime number $p$ is good for $(O, *, L, \langle \cdot, \cdot \rangle, h_0)$ if it satisfies all of the following conditions: $p$ is unramified in $O$, as in [Lan13 Def. 1.1.1.18]; $p \nmid [L^# : L]$; $p \neq 2$ whenever $(O \otimes \mathbb{R}, *)$ has a simple factor isomorphic to a matrix algebra over the Hamiltonian numbers $\mathbb{H}$ with its canonical positive involution; and there is no nontrivial level structure at $p$. See [Lan13 Def. 1.4.1.1].) Note that people often talk about good reductions instead of smoothness when describing integral models, even though it is not always clear what reductions should mean without some properness assumption.

The moduli problems above are defined using the language of isomorphism classes. If we use the language of isogeny classes instead, then it is more natural to work with rational PEL datum

$$(5.1.2.1) \quad (O \otimes \mathbb{Q}, *, L \otimes \mathbb{Q}, \langle \cdot, \cdot \rangle, h_0),$$

which is good for defining PEL moduli problems over $F_0$ (in characteristic zero) up to isogenies of all possible degrees, or with its $p$-integral version

$$(5.1.2.2) \quad (O \otimes \mathbb{Z}(p), *, L \otimes \mathbb{Z}(p), \langle \cdot, \cdot \rangle, h_0),$$

which is good for defining PEL moduli problems over $O_{F_0,(p)}$ only up to isogenies of degrees prime to $p$. The latter was done in [Kot92, Sec. 5] and generalized in [Lan13, Sec. 1.4.2], and then compared in [Lan13 Sec. 1.4.3] with the definition by isomorphism classes in [Lan13 Sec. 1.4.3]. (While the comparison was not difficult, the author still found it helpful to have the details recorded in the literature.)
5.1.3. PEL-type Shimura varieties. The PEL moduli problems as above are related to the double coset spaces in Section 2 as follows. Firstly, the subtuple \((\mathcal{O}, *, L, \langle \cdot, \cdot \rangle)\) of \((5.1.1.1)\) defines a group scheme \(G\) over \(\mathbb{Z}\) by
\[
G(R) := \{(g, r) \in \text{End}_\mathcal{O}(L \otimes R) \times R^* : \langle gx, gy \rangle = r(x, y), \text{ for all } x, y \in L \otimes R\},
\]
for each ring \(R\). The elaborate conditions for a prime \(p\) to be good, in the first paragraph of Section 5.1.2, then imply that \(G(\mathbb{Z}_p)\) is a so-called hyperspecial open compact subgroup of \(G(\mathbb{Q}_p)\) when \(G_{\mathbb{Q}_p}\) is connected as an algebraic group over \(\mathbb{Q}_p\). Secondly, the remaining entry \(h_0 : \mathbb{C}^\times \to \text{End}_\mathcal{O}\mathbb{Q}_p(L \otimes \mathbb{R})\) defines a homomorphism
\[
h_0 : \mathbb{C}^\times \to G(\mathbb{R}),
\]
whose \(G(\mathbb{R})\)-conjugacy class defines a manifold
\[
\mathcal{D} := G(\mathbb{R}) \cdot h_0.
\]
Then \((G, \mathcal{D})\) defines a double coset space \(X_\mathcal{U}\) as in \((2.2.15)\), for each neat open compact subgroup \(\mathcal{U}\) of \(G(\mathbb{A}^\infty)\). When \((G, \mathcal{D})\) is a Shimura datum (which is, however, not always the case), we say it is a PEL-type Shimura datum, and say \(X_\mathcal{U}\) is a PEL-type Shimura variety. In this case, the reflex field \(F_0\) for \(h_0\) coincides with the reflex field of \((G, \mathcal{D})\) (see Theorem 2.4.3).

Example 5.1.3.4. Suppose \(\mathcal{O} = \mathbb{Z}^\star\); \(*\) is trivial; \(L = \mathbb{Z}^{2n}\); \(\langle \cdot, \cdot \rangle\) is defined by the matrix \(2\pi i \begin{pmatrix} -1_n & 1_b \\ -1_b & 1_n \end{pmatrix}\); and \(h_0(e^{i\theta}) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}\), for all \(\theta \in \mathbb{R}\). Then \(G \cong \text{GSp}_{2n}\) (see Example 2.1.2), and \(\mathcal{D} = G(\mathbb{R}) \cdot h_0 \cong \mathcal{H}_{a,b}^\mathcal{D}\) (see (3.1.4.6)). This is the important special case of Siegel moduli problems.

Example 5.1.3.5. Suppose \(\mathcal{O} = \mathcal{O}_E\) is an order in an imaginary quadratic extension field \(E\) of \(\mathbb{Q}\), with a fixed isomorphism \(E \otimes \mathbb{R} \cong \mathbb{C}; \(*\) is the complex conjugation of \(E\) over \(\mathbb{Q}\); \(L = \mathcal{O}^{a+b}\) for some integers \(a \geq b \geq 0\); \(\langle \cdot, \cdot \rangle\) is defined by \(2\pi i\) times the composition of the skew-Hermitian pairing defined by some matrix
\[
\begin{pmatrix} \varepsilon 1_{a-b} & 1_b \\ -1_b & 1_n \end{pmatrix}
\]
with \(\text{tr}_{\mathcal{O}/\mathbb{Z}} : \mathcal{O} \to \mathbb{Z}\), where \(\varepsilon \in \mathcal{O}\) satisfies \(-\varepsilon \in \mathbb{R}_{>0}\); and \(h_0(e^{i\theta}) = \begin{pmatrix} \cos \theta & e^{-i\theta} \\ \sin \theta & \cos \theta \end{pmatrix}\), for all \(\theta \in \mathbb{R}\). Then \(G(\mathbb{R}) \cong \text{GU}_{a,b} \cong \text{GU}'_{a,b}\); and \(\mathcal{D} = G(\mathbb{R}) \cdot h_0 \cong \mathcal{H}_{a,b}^{\mathcal{D}}\) (see (3.2.5.9)).

Example 5.1.3.6. Suppose \(\mathcal{O}\) is an order in a quaternion algebra \(B\) over \(\mathbb{Q}\) such that \(B \otimes \mathbb{R} \cong \mathbb{H}\) and \(\mathcal{O}\) is stabilized by the canonical positive involution \(*\); \(L = \mathcal{O}^{2n}\); \(\langle \cdot, \cdot \rangle\) is \(2\pi i\) times the composition of the skew-Hermitian pairing defined by the matrix \(\begin{pmatrix} -1_n & 1_b \\ -1_b & 1_n \end{pmatrix}\) with \(\text{tr}_{\mathcal{O}/\mathbb{Z}} : \mathcal{O} \to \mathbb{Z}\); and \(h_0(e^{i\theta}) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}\), for all \(\theta \in \mathbb{R}\). Then the derived group of \(G(\mathbb{R})\) is isomorphic to \(\text{SO}^*_{2n}\), and \(\mathcal{D} = G(\mathbb{R}) \cdot h_0\) contains \(\mathcal{H}_{\text{SO}^*_{2n}}\) as a connected component (see Section 3.3.1). However, \(G\) is not connected as an algebraic group over \(\mathbb{Q}\).

These three examples are prototypical for integral PEL data and the associated PEL moduli problems. By the classification of positive involutions (see the review
in \cite{Lan13} Prop. 1.2.1.13 and 1.2.1.14] based on Albert’s classification, or just the classification over \( \mathbb{R} \) in \cite{Kot92} Sec. 2), up to cocenter (i.e., up to replacing the group with a subgroup containing the derived group), \( G(\mathbb{R}) \) factorizes as a product of groups of the form \( \text{Sp}_{2n}(\mathbb{R}), \text{U}_{a,b}, \) or \( \text{SO}^*_n \) (with varying values of \( n \geq 0 \) and \( a \geq b \geq 0 \)). Thus, the connected components of \( D \) are products of the Hermitian symmetric domains we have seen in Sections 3.1, 3.2, and 3.3. However, we never see symmetric domains we have seen in Sections 3.1, 3.2, and 3.3. However, we never see

\[ a \geq b \geq 0 \]

\[ (5.1.3.7) \]

\[ X \rightarrow M_\mathcal{U}(\mathbb{C}) \]

(5.1.3.7) which is generally not an isomorphism, due to the so-called failure of Hasse’s principle (see \cite{Lan13} Rem. 1.4.3.12 and \cite{Kot92} Sec. 7–8). More concretely, this is because the definition of \( M_\mathcal{U} \) uses only the adelic group \( G(\mathbb{A}) \)—the level structures use \( G(\mathbb{A}^\infty) \), while the Lie algebra condition uses \( G(\mathbb{R}) \). But there can be more than one mutually nonisomorphic algebraic groups \( G^{(1)}, \ldots, G^{(m)} \) (including \( G \)) over \( \mathbb{Q} \) such that \( G^{(j)}(\mathbb{A}) \cong G(\mathbb{A}) \), for all \( j \), and such that \( M_\mathcal{U}(\mathbb{C}) \) is the disjoint union of the \( X^{(j)}_\mathcal{U} \)’s defined using \( G^{(j)} \)’s. (When the pairs \((G^{(j)}, D)\)’s are Shimura data, by checking the action of \( \text{Aut}(\mathcal{U}/F_0) \) on the dense subset of the so-called CM points, it can be shown that this disjoint union descends to a disjoint union over \( F_0 \) of the canonical models of the \( X^{(j)}_\mathcal{U} \)’s.) In cases of types A and C, such a failure is harmless in practice, because all the \( X^{(j)}_\mathcal{U} \)’s are isomorphic to each other, even at the level of canonical models over \( F_0 \) (see \cite{Kot92} Sec. 8). However, in cases of type D, such a failure can be quite serious. It can even happen that some of the \( X^{(j)}_\mathcal{U} \)’s are noncompact, while others are not (see \cite{Lan13} Ex. A.7.2).

With the tacit running assumption that \( \mathcal{U} \) is neat (see \cite{Pin89} 0.6) again, or see \cite{Lan13} Def. 1.4.1.8), \( M_\mathcal{U} \) is a (smooth) quasi-projective scheme because it is finite
over the base change of some Siegel moduli scheme (see Section 3.1.5), and hence an open and closed subscheme of \( \mathcal{M}_U \) serves as an integral model of \( X_U \). (When \( U \) is not neat, we still obtain a Deligne–Mumford stack whose associated coarse moduli space is a quasi-projective scheme.)

Remark 5.1.3.8. The consideration of certain Shimura varieties or connected Shimura varieties as parameter spaces for complex abelian varieties with PEL structures has a long history—see, for example, Shimura’s articles [Shi63b] and [Shi66]. The theory over the complex numbers is already quite complicated and heavy in notation, which seems inevitable for anything involving semisimple algebras and positive involutions, but there is an additional subtlety when formulating the PEL structures over the integers. That is, we can no longer talk about the singular homology or cohomology of an abelian variety over a general base field, let alone their relative versions for an abelian scheme over a general base scheme. Instead, we have resorted to the (relative) de Rham and étale homology or cohomology. This is the reason that PEL moduli problems suffer from the failure of Hasse’s principle, and that these moduli problems can only be easily defined in mixed characteristics \((0,p)\) when there is no ramification or level at \( p \). (In Section 5.1.5 below, we will summarize the known constructions when there are indeed some ramification and levels at \( p \).)

5.1.4. Toroidal and minimal compactifications of PEL moduli. The above PEL moduli \( \mathcal{M}_U \) carry universal or tautological objects, which are abelian schemes with additional PEL structures, essentially by definition.

In the important special case of Siegel moduli problems considered in Example 5.1.3.4, the tautological objects are just principally polarized abelian schemes with level structures (but with no nontrivial endomorphism structures or polarization degrees specified in the moduli problem). By vastly generalizing results due to Mumford (see [Mum72]), Faltings and Chai constructed toroidal compactifications of the Siegel moduli problems by studying the degeneration of polarized abelian schemes into semi-abelian schemes (see [FC90, Ch. III and IV]). Over such toroidal compactifications, they can algebro-geometrically define modular forms, and construct minimal compactifications of Siegel moduli problems as the projective spectra of some graded algebras of modular forms (with inputs of positivity results from [MB85]; see [FC90, Ch. V]). The main difficulty in Faltings and Chai’s construction was that they needed to glue together boundary strata parameterizing degenerations of different ranks. This was much more complicated than the earlier boundary construction in [Rap78], based on the original results of [Mum72].

The generalization of the results of [FC90, Ch. III–V] to all smooth PEL moduli problems was carried out in [Lan13]. Later in [Lan12a], it was shown that the toroidal and minimal compactifications in [Lan13] (and also in [FC90, Ch. III–V]), both based on the theory of degeneration, are indeed compatible with the toroidal and minimal compactifications in [Har89] and [Pin89] (based on the constructions in [BL89] and [AMRT75] using analytic coordinate charts). In other words, toroidal and minimal compactifications of PEL moduli problems provide integral models of toroidal and minimal compactifications of PEL-type Shimura varieties.

Over the integral models of PEL-type Shimura varieties and their toroidal compactifications, by considering the relative cohomology of certain Kuga families (which are abelian scheme torsors which are integral models of certain mixed Shimura varieties as in Section 4.2.8) and their toroidal compactifications, we
can also define automorphic vector bundles and their canonical and subcanonical extensions over these integral models (see [FC90 Ch. VI], [Lan12c especially Sec. 6], [LS12 Sec. 1–3], and [LS13 Sec. 4–5]), whose global sections and cohomology classes then provide an algebro-geometric definition of modular forms in mixed characteristics, generalizing the theory over \( \mathbb{C} \) in Section 4.2.7. See [Lan12b] and [Lan19] again for some surveys on this topic. See also [Hid02] and [Hid04], for example, for some application to the theory of \( p \)-adic modular forms, with the assumptions on the integral models of toroidal and minimal compactifications justified by the constructions in [Lan13].

5.1.5. Integral models not defined by smooth moduli. Let \( p \) be any prime number, which is not necessarily good for the integral PEL datum. (The conditions for being good for the integral PEL datum are the same as the conditions for being good for defining the PEL moduli, except that we ignore the condition of having no nontrivial level structure at \( p \).)

Let \( \hat{\mathbb{Z}}_p := \lim_{\leftarrow} \mathbb{Z}/N\mathbb{Z} \) and \( \mathbb{A}^\infty_p := \hat{\mathbb{Z}}_p \otimes \mathbb{Q} \), with the superscript “\( p \)” meaning “away from \( p \).” When the level \( \mathcal{U} \subset G(\mathbb{A}^\infty) \) is of the form \( \mathcal{U} = \mathcal{U}^p \mathcal{U}_p \), where \( \mathcal{U}^p \subset G(\mathbb{A}^\infty_p) \) and \( \mathcal{U}_p \subset G(\mathbb{Q}_p) \) such that \( \mathcal{U}_p \) is a parahoric open compact subgroup of \( G(\mathbb{Q}_p) \)—that is, when \( \mathcal{U}_p \) is the identity component of the stabilizer of some multichain of \( \mathbb{Z}_p \)-lattices in \( L \otimes \mathbb{Q}_p \)—there is also a natural way to define a PEL moduli problem parameterizing not just individual abelian schemes with additional PEL structures, but rather multichains of isogenies of them. (See [RZ96 Ch. 3 and 6].) This generalizes the idea that, in the case of modular curves, the moduli problem at level \( \Gamma_0(p) \) parameterizes degree-\( p \) cyclic isogenies of elliptic curves. (See, for example, [KM85 Sec. 3.4 and 6.6], which gave a moduli interpretation of the corresponding construction in [DR73] that was by merely taking normalizations.)

PEL moduli problems at parahoric levels at \( p \) are generally far from smooth in mixed characteristics \( (0, p) \), in which case we say that the integral models have bad reductions. Nevertheless, people have developed a theory of local models which describes the singularities of such moduli problems—see, for example, [RZ96], [PR03], [PR07], [PR09], and [PZ13]. (See also [Rap05], [Hai05], and [PRS13] for some surveys.) Beginners should be warned that the forgetful morphisms from a higher parahoric level to a lower one, or more precisely from one such moduli problem defined by some multichain of \( \mathbb{Z}_p \)-lattices to another one defined by a simpler multichain, is usually just projective but not finite. This is very different from what we have seen in characteristic zero (or in the classical modular curve case in [DR73] and [KMS]).

For higher levels at \( p \), the older ideas in [DR73] by taking normalizations, and in [KMS] by generalizing the so-called Drinfeld level structures, are both useful. But they are useful in rather different contexts—exactly which strategy is more useful depends on the intended applications.

The idea of generalizing Drinfeld level structures played important roles in [HT01], [Man04], [Man05], and [Man11], and in subsequent works based on them such as [Shi11], where people needed to compute the cohomology of certain nearby cycles and count points, and hence would prefer to work with moduli problems even when they are not flat. (It was known much earlier—see [CN90 Appendix A]—that moduli problems defined by Drinfeld level structures are generally not even flat over the base rings.) But such moduli problems tend to be very
complicated, and might be too cumbersome when it comes to the construction of compactifications, because most known techniques of gluing require at least some noetherian normal base schemes. (The results in [HT01], [Man04], [Man05], [Man11], and [Shi11] were all for compact PEL-type Shimura varieties.)

On the contrary, the idea of taking normalizations produces, essentially by definition, noetherian normal schemes flat over the base rings. It had been overlooked for a long time because of its lack of moduli interpretations, and because no theory of local models was available for them. While these two problems still remain, recent developments beyond the PEL-type cases (see the discussion on the Hodge-type cases below in Section 5.2.3), and also developments in the construction of $p$-adic modular forms (as in [HLTT16], which was based on [Lan18a] and did not require the integral models to be constructed as moduli problems) motivated people to systematically reconsider the constructions by taking normalizations. In the PEL-type cases, we can construct good compactifications for all such integral models (and their variants)—see [Lan16], [Lan17], and [Lan18b]. Such compactifications are good in the sense that the complications in describing its local structures and its boundary stratifications are completely independent of each other. This is useful in, for example, [LS18a] Sec. 6.3, which generalized the results of [Man04] and [Man11] to all noncompact cases, without having to compactify any moduli problems defined by Drinfeld level structures.

In [LS18a], it is shown that many familiar subsets of the reductions of the integral models of Shimura varieties in characteristic $p > 0$, such as $p$-rank strata, Newton strata, Oort central leaves, Ekedahl–Oort strata, and Kottwitz–Rapoport strata, and their pullbacks to the integral models defined by taking normalizations at arbitrarily higher levels at $p$, admit partial toroidal and minimal compactifications with many properties similar to the whole toroidal and minimal compactifications.

5.1.6. Langlands–Rapoport conjecture. In [Lan76] and [Lan77], where the terminology Shimura varieties first appeared in the literature, Langlands proposed to study Hasse–Weil zeta functions of Shimura varieties and relate them to the automorphic $L$-functions, generalizing earlier works for $GL_2$ and its inner forms by Eichler, Shimura, Kuga, Sato, and Iitaka. This is very ambitious—see the table of contents of [LR92] for the types of difficulties that already showed up in the case of Picard modular surfaces. This motivated the Langlands–Rapoport conjecture, which gives a detailed description of the points of the reduction modulo $p$ of a general Shimura variety at a hyperspecial level at $p$—see [LR87], [Pla96], [Rei97], and [Rap05]. See also [Kot90] for a variant of the conjecture due to Kottwitz, which was proved by himself for all PEL-type Shimura varieties of types A and C in [Kot92].

5.2. Hodge-type and abelian-type cases.

5.2.1. Hodge-type Shimura varieties.

**Definition 5.2.1.1.** A Hodge-type Shimura datum $(G, D)$ is a Shimura datum such that there exists a Siegel embedding

$$\text{(5.2.1.2)} \quad (G, D) \hookrightarrow (\text{GSp}_{2n}, H_n^\pm);$$

namely, an injective homomorphism $G \hookrightarrow \text{GSp}_{2n}$ of algebraic groups over $\mathbb{Q}$ inducing an (automatically closed) embedding $D \hookrightarrow H_n^\pm$, for some $n \geq 0$ (cf. Remarks 2.4.11 and 2.4.12). A Hodge-type Shimura variety is a Shimura variety associated with a Hodge-type Shimura datum.
If \( D = G(\mathbb{R}) \cdot h_0 : \mathbb{C}^\times \to G(\mathbb{R}) \) as before, then the above means the composition of \( h_0 \) with \( G(\mathbb{R}) \to \text{GSp}_n(\mathbb{R}) \) as in (3.1.4.7). As explained in Remark 2.4.13 each neat open compact subgroup \( \mathcal{U} \) of \( G(\mathbb{A}_\infty) \) is the pullback of some neat open compact subgroup \( \mathcal{U}' \) of \( \text{GSp}_n(\mathbb{A}_\infty) \) such that the Shimura variety \( X_{\mathcal{U}} \) associated with \((G, D)\) at level \( \mathcal{U} \) is a special subvariety of the Siegel modular variety at level \( \mathcal{U}' \). Then the points of \( X_{\mathcal{U}} \) are equipped with polarized abelian varieties with level structures, which are pulled back from the Siegel modular varieties (and hence depend on the choice of the Siegel embedding above).

Example 5.2.1.3. All the PEL-type Shimura data are, essentially by definition, of Hodge type, because of the definition of \( G \) in (5.1.3.1), and because of the condition on \( h_0 \) in Section 5.1.1. But there are many Hodge-type Shimura data which are not of PEL type. Important examples of this kind are given by general spin groups (with associated Hermitian symmetric domains as in Section 3.4.3), which admit Siegel embeddings thanks to the so-called Kuga–Satake construction (see [Sat66], [Sat68], and [KS67]; see also the reviews in [Del72, Sec. 3–4] and [MP16, Sec. 1 and 3]).

The injectivity of the homomorphism \( G \to \text{GSp}_n \) means we have a faithful \( 2n \)-dimensional symplectic representation of \( G \). By [DMOSS2, Ch. I, Prop. 3.1], the faithfulness of this representation implies that \( G(\mathbb{Q}) \) is the subgroup of \( \text{GSp}_n(\mathbb{Q}) \) fixing some tensors \( (s_i) \) in the tensor algebra generated by \( \mathbb{Q}^{2n} \) and its dual. But since \( \mathbb{C}^\times \xrightarrow{h_0} G(\mathbb{R}) \to \text{GSp}_n(\mathbb{R}) \) defines a Hodge structure on \( \mathbb{R}^{2n} \), by [DMOSS2, Ch. I, Prop. 3.4], these tensors \( (s_i) \) are Hodge tensors; namely, tensors of weight 0 and Hodge type \((0, 0)\) with respect to the induced Hodge structure on the tensor algebra of \( \mathbb{R}^{2n} \).

Then the Shimura varieties \( X_{\mathcal{U}} \) associated with \((G, D)\) can be interpreted as parameter spaces of polarized abelian varieties (over \( \mathbb{C} \)) equipped with some Hodge cycles and level structures. According to Deligne (see [DMOSS2, Ch. I, Main Thm. 2.11]), Hodge cycles on complex abelian varieties are absolute Hodge cycles. Roughly speaking, this means that the Hodge cycles can be conjugated, as if they were known to be defined by algebraic cycles (as if the Hodge conjecture were true for them), when the underlying abelian varieties are conjugated by an automorphism of \( \mathbb{C} \). This defines a canonical action of \( \text{Aut}(\mathbb{C}/F_0) \) on the points of \( X_{\mathcal{U}} \) by conjugating the Hodge cycles of the corresponding abelian varieties, which coincides with the action given by the structure of \( X_{\mathcal{U}} \) as the \( \mathbb{C} \)-points of its canonical model over \( F_0 \). Because of this, people can formulate some moduli interpretation for the points of the canonical model of \( X_{\mathcal{U}} \) over \( F_0 \).

Such a pointwise moduli interpretation has not really helped people write down any useful moduli problem, but they are still important for studying the less direct construction by taking normalizations, to be introduced in Section 5.2.3 below.

5.2.2. Abelian-type Shimura varieties.

Definition 5.2.2.1. An abelian-type Shimura datum \((G, D)\) is a Shimura datum such that there exist a Hodge-type Shimura datum \((G_1, D_1)\) and a central isogeny
\[
G_1^{\text{der}} \to G^{\text{der}}
\]
between the derived groups which induces an isomorphism

\[(G_{1}^{\text{ad}}, D_{1}^{\text{ad}}) \sim (G^{\text{ad}}, D^{\text{ad}})\]

between the adjoint quotients. An abelian-type Shimura variety is a Shimura variety associated with an abelian-type Shimura datum.

By definition, every Hodge-type Shimura datum (resp. Shimura variety) is of abelian type. Essentially by definition, the connected components of an abelian-type Shimura variety are finite quotients of those of some Hodge-type Shimura variety. Thus, results for abelian-type Shimura varieties are often proved by reducing to the case of Hodge-type Shimura variety.

Example 5.2.2.4. The simplest and perhaps the most prominent examples of abelian-type Shimura varieties that are not of Hodge type are the Shimura curves, which are associated with Shimura data \((G, D)\) with \(G(\mathbb{Q}) = B^{\times}\), where \(B\) is a quaternion algebra over a totally real extension \(F\) of \(\mathbb{Q}\) of degree \(d > 1\) such that

\[B \otimes \mathbb{Q} \cong M_{2}(\mathbb{R}) \times \mathbb{H}^{d-1},\]

in which case

\[D \cong H^{\pm}\]

is one-dimensional. More generally, there are abelian-type Shimura data \((G, D)\) with \(G(\mathbb{Q}) = B^{\times}\), where \(B\) is a quaternion algebra over \(F\) such that

\[B \otimes \mathbb{Q} \cong M_{2}(\mathbb{R})^{a} \times \mathbb{H}^{b}\]

for some \(a, b \geq 1\), in which case

\[D \cong (H^{\pm})^{a}\]

(cf. [Shi71, Sec. 9.2]). Even more generally, as explained in [Del71b, Sec. 6], with the same \(B\) and \(F\) (and \(a, b \geq 1\)), for each integer \(n \geq 1\), there are abelian-type Shimura data \((G, D)\) such that \(G(\mathbb{Q})\) is the group of similitudes over \(F\) of a symmetric bilinear pairing over \(B^{n}\), and such that

\[G(\mathbb{R}) \cong \text{GSp}_{2n}(\mathbb{R})^{a} \times \mathbb{H}^{b},\]

where \(H\) is a real form of \(\text{GSp}_{2n}(\mathbb{C})\) with compact derived group (which is \(\mathbb{H}^{n}\) when \(n = 1\)), in which case

\[D \cong (H^{\pm})^{a}\]

In all these cases, the related Hodge-type Shimura data (as in Definition 5.2.2.1) and their Siegel embeddings (as in Definition 5.2.1.1) can be constructed with some auxiliary choice of imaginary quadratic field extensions of \(\mathbb{Q}\) (see [Del71b, Sec. 6], and also [Car86, Sec. 2]). However, none of these are of Hodge type (let alone of PEL type) because the Hodge structures defined by rational representations of \(G\) are not rational Hodge structures (see [Mil05, Sec. 9, part (c) of the last example]).

Example 5.2.2.11. There are also the abelian-type Shimura varieties associated with Shimura data \((G, D)\) such that the derived group of \(G\) as an algebraic group over \(\mathbb{Q}\) is the special orthogonal group defined by some symmetric bilinear form over \(\mathbb{Q}\) of signature \((a, 2)\) over \(\mathbb{R}\), for some \(a > 1\), so that \(G(\mathbb{R})^{+}\) is up to cocenter the group \(\text{SO}_{a,2}(\mathbb{R})^{+}\) studied in Section 3.4.3 (See also the last paragraph of Section 3.4.4). In this case, the related Hodge-type Shimura data (as in Definition 5.2.2.1) are
general spin groups, with their Siegel embeddings (as in Definition 5.2.1.1) given
by the Kuga–Satake construction, as in Example 5.2.1.3 mapping G into GSp\(_n\) with \(n = 2^\alpha\) (see the same references in Example 5.2.1.3). Such Shimura varieties
associated with special orthogonal groups are important because they are related
to the moduli of interesting algebraic varieties. For example, when \(\alpha = 19\) in the
above, the corresponding Shimura varieties contain as open subspaces the moduli
spaces of polarized K3 surfaces with level structures (see, for example, \[Riz06\] Sec.
6], [MP15 Sec. 2 and 4], and \[She17\] Sec. 4.2).

In \[Del79\] 2.3, based on earlier works due to Satake (see \[Sat65\]; and see also
\[Sat67\] and \[Sat80\] Ch. IV], Deligne analyzed all abelian-type Shimura data \((G, D)\)
that are adjoint and simple in the sense that \(G\) is adjoint and simple as an algebraic
group over \(\mathbb{Q}\). (See also \[Mil13\] Sec. 10.) When \(G\) is simple over \(\mathbb{Q}\), there is up
to isomorphism a unique irreducible Dynkin diagram associated with all the simple
factors of \(G_\mathbb{C}\), or rather \(\text{Lie} \, G(\mathbb{C})\). (In fact, in this case, \(G_\mathbb{ad} \cong \text{Res}_{k/\mathbb{Q}} \, H\) for some
number field \(k\) and some connected semisimple algebraic group \(H\) over \(k\) such that \(H_{\overline{\mathbb{Q}}}\)
remains simple over the algebraic closure \(\overline{\mathbb{Q}}\) of \(\mathbb{Q}\) in \(\mathbb{C}\); see,
for example, \[BT65\] 6.21(ii)]). We say that the group is of type \(A_n\), \(B_n\), \(C_n\), \(D_n\),
\(E_6\), or \(E_7\) accordingly to the type of such a Dynkin diagram (cf. the summary in
Section 3.7). In cases of type \(D_n\), for \(n \geq 5\), we further single out the special case of
\(D_{nR}\) (resp. \(D_{nH}\)) if all the noncompact simple factors of \(\text{Lie} \, G(\mathbb{R})\) are isomorphic
to \(\text{so}_{2n-2}\) (resp. \(\text{so}_{n}^*\)). In cases of type \(D_4\), the situation is more complicated
because \(\text{so}_{5,2} \cong \text{so}_6^*\), but we can still single out some special cases of type \(D_4^R\)
and \(D_4^H\). For simplicity, we shall drop the subscript \(n\) when it is not used in the discussion.
Deligne showed that an adjoint and simple Shimura datum \((G, D)\) is of
abelian type exactly when it is of types \(A, B, C, D_{nR}\) and \(D_{nH}\). In cases of types \(A, B, C, \) and \(D^R\), the related Hodge-type Shimura datum \((G_1, D_1)\) (as in Definition
5.2.1.1) can be chosen such that the derived group \(G_1^{\text{der}}\) is simply-connected as an
algebraic group over \(\mathbb{Q}\). However, in the remaining cases of \(D_{nH}\), where \(n \geq 4\), the
related Hodge-type Shimura datum can only be chosen such that the noncompact simple factors of \(G_1^{\text{der}}(\mathbb{R})\) are all isomorphic to \(\text{SO}_{2n}^*\). (Although this could have been a more precise statement in terms of algebraic groups, we are content with
the simpler statement here in terms of real Lie groups.)

More generally, suppose that \((G, D)\) is a Shimura datum that is simple in the
sense that the adjoint quotient \(G_\mathbb{ad}\) is simple. (But we no longer assume that \(G\)
itself is adjoint.) Then we can still classify the type of \((G, D)\) by classifying the type of \((G_\mathbb{ad}, D_\mathbb{ad})\) as above, and we have the following possibilities:

1. All cases of types \(A, B, C, \) and \(D^R\) are of abelian type.
2. Cases of type \(D^H\) can be of abelian type only when \(G_\mathbb{der}\) is the quotient of
\(G_1^{\text{der}}\) for the Hodge-type datum \((G_1, D_1)\) mentioned in the last paragraph.
In particular, \(G_1^{\text{der}}(\mathbb{R})\) cannot have factors isomorphic to the spin group
cover \(\text{Spin}_{2n}\) of \(\text{SO}_{2n}^*\).
3. Cases of mixed type \(D\), namely, those that are neither of type \(D^R\) nor of
\(D^H\), are not of abelian type. Such cases exist by a general result due
to Borel and Harder (see \[BH78\] Thm. B)).
4. Cases of type \(E_6\) and \(E_7\) are never of abelian type.

Thus, roughly speaking, all Shimura varieties associated with symplectic and unitary
groups are of abelian type, and none of the Shimura varieties associated with
exceptional groups can be of abelian type, while there are some complications for Shimura varieties associated with spin groups or special orthogonal groups.

5.2.3. Integral models of these Shimura varieties. The strategy for constructing integral models of Hodge-type Shimura varieties is as follows. Suppose \((G, D)\) is a Hodge-type Shimura datum with Siegel embedding

\[
(G, D) \hookrightarrow (\text{GSp}_n, \mathcal{H}^\pm_n),
\]

and suppose \(U\) is an open compact subgroup of \(G(\mathbb{A}^\infty)\) that is the pullback of some open compact subgroup \(U'\) of \(\text{GSp}_{2n}(\mathbb{A}^\infty)\) such that \(U' = U'^{p}U'_p\), where \(U'^{p}\) is some neat open compact subgroup of \(\text{GSp}_{2n}(\mathbb{A}^{\infty, p})\), and where \(U'_p = \text{GSp}_{2n}(\mathbb{Z}_p)\).

In the literature, people have considered the more general setting where \(U'_p\) is the stabilizer in \(\text{GSp}_{2n}(\mathbb{Q}_p)\) of some \(\mathbb{Z}_p\)-lattice, which can nevertheless be reduced to our setting here by modifying the Siegel embedding \((5.2.3.1)\), at the expense of replacing \(n\) with \(8n\), using “Zarhin’s trick” as in \([\text{Lan}16\text{, Lem. 4.9]}\) or \([\text{Lan}18a\text{, Lem. 2.1.1.9]}\). Let \(v_p\) be a place of \(F_0\), and let \(O_{F_0,(v)}\) denote the localization of \(F_0\) at \(v\). Then we have a canonical morphism

\[
X_{U,F_0} \to X_{U',O_{F_0,(v)}},
\]

where \(X_{U,F_0}\) abusively denotes the canonical model of \(X_{U}\) over \(F_0\), and where \(X_{U',O_{F_0,(v)}}\) denotes the base change to \(O_{F_0,(v)}\) of the Siegel moduli scheme at level \(U'\). Then we define the integral model

\[
X_{U,O_{F_0,(v)}}
\]

of \(X_{U,F_0}\) over \(O_{F_0,(v)}\) by taking the normalization of \(X_{U',O_{F_0,(v)}}\) in \(X_{U,F_0}\) under the above canonical morphism.

A priori, this is a very general construction, which also makes sense when \(U\) is just a subgroup of the pullback of \(U'\). This is what we did in the PEL-type cases in \([\text{Lan}18a]\) and \([\text{Lan}16]\), as reviewed in Section 5.1.3, and similar ideas can be traced back to \([\text{DR}73]\).

But when \(U\) is of the form \(U'U_p\) such that \(U'^{p}\) is a neat open compact subgroup of \(G(\mathbb{A}^{\infty, p})\) and such that \(U_p\) is a hyperspecial maximal open compact subgroup of \(G(\mathbb{Q}_p)\), Vasiu, Kisin, and others (see \([\text{Mil}92\text{, Rem. 2.15}], [\text{Vas}99]\), \([\text{Moo}98]\), \([\text{Kis}10]\), and \([\text{KMP}16]\) proved that \(X_{U,O_{F_0,(v)}}\) is smooth and satisfies certain extension properties making it an integral canonical model. (The formulations of the extension properties vary from author to author. See, in particular, \([\text{Mil}92\text{, Sec. 2}], [\text{Moo}98\text{, Sec. 3}], [\text{Kis}10\text{ (2.3.7)}]\). An integral canonical model then just means an integral model satisfying some formulation of the extension property. This notion is not expected to be useful when the levels at \(p\) are not hyperspecial.) They also showed that, roughly speaking, by working with connected components, by taking suitable quotients of integral models of Hodge-type ones by finite groups, and by descent, we also obtain integral models for all abelian-type Shimura varieties at hyperspecial levels at \(p\), which are smooth and satisfy the same extension properties.

When \(p > 2\), and when \(G\) splits over a tamely ramified extension of \(\mathbb{Q}_p\), Kisin and Pappas proved in \([\text{KP}18]\) that all abelian-type Shimura varieties at parahoric levels at \(p\) have integral models which have local models as predicted by the group-theoretic construction in \([\text{PZ}13]\), and which satisfy a weaker extension property.
In [Kis17], Kisin proved a variant of the Langlands–Rapoport conjecture (partly following Kottwitz’s variant; cf. Section 5.1.6) for all abelian-type Shimura varieties at hyperspecial levels at \( p \), when \( p > 2 \).

### 5.2.4. Integral models of toroidal and minimal compactifications.

In [MP19], it is proved that, by taking normalizations over the toroidal and minimal compactifications of Siegel moduli constructed in [FC90] and [Lan13], we also have good integral models of toroidal and minimal compactifications of integral models of Hodge-type Shimura varieties constructed in the previous subsection. The cone decompositions for the toroidal compactifications thus obtained are rather inexplicit pullbacks of certain smooth ones for the Siegel moduli. But it is believed that, by considering normalizations of blowups of the minimal compactifications as in [Lan17], we can obtain toroidal compactifications associated with more general projective and smooth cone decompositions as well.

As in the PEL-type cases in Section 5.1.5, such compactifications are good in the sense that the complications in describing its local structures and its boundary stratifications are completely independent of each other. Using the toroidal compactifications thus constructed, we showed in [LS18b] and [LS18c] that, for each integral model of Hodge-type or abelian-type Shimura varieties constructed in [Kis10], [KMP16], and [KP18], and for any prime \( \ell \) different from the residue characteristic, the \( \ell \)-adic étale nearby cycle cohomology over the special fiber is canonically isomorphic to the \( \ell \)-adic étale cohomology over the (geometric) generic fiber, without resorting to the proper base change theorem (cf. [AGV73, XII, 5.1] and [DK73a, XIII, (2.1.7.1) and (2.1.7.3)]). (In [LS18c], the proof in the case of abelian-type Shimura varieties is achieved by reduction to the case of Hodge-type Shimura varieties.) Intuitively speaking, the special fibers of these integral models have as many points as there should be, at least for studying the \( \ell \)-adic étale cohomology of Shimura varieties.

### 5.3. Beyond abelian-type cases.

Although abelian-type Shimura varieties are already very useful and general, as we have seen in Section 5.2.2, they do not cover all possibilities of Shimura varieties. Among those associated with simple Shimura data, we are not just missing those of types \( E_6 \) and \( E_7 \)—we are also missing the majority (depending on one’s viewpoint) of type D cases. But these missing cases can still be quite interesting, both for geometric and arithmetic reasons.

For example, in [MS10], Mihăeş and Suh showed that there are many examples of geometrically connected quasi-projective varieties \( X \) over number fields \( F \) such that the complex analytifications of \( X \otimes \mathbb{C} \) and \( X \otimes \mathbb{C} \) have nonisomorphic fundamental groups for different embeddings \( \sigma \) and \( \tau \) of \( F \) into \( \mathbb{C} \), by considering connected Shimura varieties of mixed type D (which exist by [BH78, Thm. B]).

As another example, we have the following result of R. Liu and X. Zhu’s:

**Theorem 5.3.1** (see [LZ17, Thm. 1.2]). Suppose \( X_{\mathcal{U}} \) is a Shimura variety associated with a general Shimura datum \( (G, D) \) (which can certainly be simple of types \( E_6 \), \( E_7 \), or mixed D) at some level \( \mathcal{U} \). Suppose \( \mathcal{V} \) is an étale \( \mathbb{Q}_p \)-local system associated with some finite-dimensional representation of \( G^\circ = G/Z_s(G) \), where \( Z_s(G) \) is the minimal subtorus \( Z_s(G) \) of the center \( Z(G) \) such that the torus \( Z(G)^\circ/Z_s(G) \) has the same split ranks over \( \mathbb{Q} \) and \( \mathbb{R} \). Suppose \( x : \text{Spec}(\overline{F}) \to X_{\mathcal{U}} \) is a closed point of \( X_{\mathcal{U}} \), together with a geometric point \( \bar{x} : \text{Spec}(\overline{F}) \to X_{\mathcal{U}} \) lifting \( x \). Then the stalk

\[ \lim_{\text{pro-\'etale}} \pi^{-1}(x) \]
\( \mathcal{V}_F \), when regarded as a representation of \( \text{Gal}(\overline{F}/F) \), is \textbf{geometric} in the sense of Fontaine–Mazur (see [FM97, Part I, Sec. 1]). That is, it is unramified at all but finitely many places of \( F \), and is de Rham at the places of \( F \) above \( p \).

Recall that Fontaine and Mazur conjectured that every irreducible geometric Galois representation \textit{comes from geometry} in the sense that it appears as a subquotient of the \( \acute{e} \text{tale} \) cohomology of an algebraic variety (see [FM97, Part I, Sec. 1] again). Theorem 5.3.1 would have been less surprising if it only considered abelian-type Shimura varieties with certain assumptions such that the stalks as in the theorem are just the \( p \)-adic realizations of some \textit{abelian motives}. But the theorem does consider non-abelian-type Shimura varieties as well, and we do not know any families of motives parameterized by these general Shimura varieties. We hope that there are such families, but this is a difficult problem that is still wide open.

Both results mentioned above are about Shimura varieties or connected Shimura varieties in characteristic zero. But what can we say about their integral models? Certainly, any quasi-projective variety over the reflex field \( F_0 \) has some integral model over the ring of integers \( \mathcal{O}_{F_0} \) in \( F_0 \), but we are not interested in such a general construction. It would be desirable to have a construction such that the pullback to \( \mathcal{O}_{F_0,(p)} \) of the integral model at level \( U \) is smooth and satisfies some reasonable extension property, when \( U \) is of the form \( U^0 U_p \), such that \( U_p \) is a neat open compact subgroup of \( G(A^\infty,p) \) and such that \( U_p \) is a \textit{hyperspecial} maximal open compact subgroup of \( G(Q_p) \). Moreover, it would be desirable that the reduction modulo \( p \) of such integral models can be described by some variant of the Langlands–Rapoport conjecture (see Section 5.1.6 and the end of Section 5.2.3). These are also difficult and wide open problems, but their nature might be very different from that of the problem (in the previous paragraph) of finding motives parameterized by such Shimura varieties.

5.4. Beyond Shimura varieties? There are some exciting recent developments which are not about integral models of Shimura varieties, or not even about Shimura varieties, but it would have been a shame not to mention them at all.

In [Sch15], Scholze constructed \textit{perfectoid Shimura varieties} in all Hodge-type cases, which are limits at the infinite level at \( p \) (in a subtle sense) of Shimura varieties at finite levels. Base on these, in [She17], Shen also constructed perfectoid Shimura varieties in all abelian-type cases. Although the constructions in [Sch15] used Siegel moduli schemes over the integers, these perfectoid Shimura varieties are objects (called \textit{adic spaces}) defined over the characteristic zero base field \( \mathbb{C}_p \) (the completion of an algebraic closure \( \overline{\mathbb{Q}}_p \) of \( \mathbb{Q}_p \)). But the property of being \textit{perfectoid} (see [Sch12]) makes them much more powerful than integral models at finite levels for studying many questions about torsion in the cohomology of Shimura varieties and related geometric objects—see, for example, [Sch15], [NT16], and [CS17].

Even the concept of Shimura varieties itself has been challenged and generalized. In his Berkeley–MSRI lectures in the Fall of 2014 (see [SW20]), Scholze defined certain moduli spaces of \textit{mixed-characteristic shtukas}, which can be viewed as a local analogue of Shimura varieties, but is much more general than the previously studied local geometric objects such as the so-called \textit{Rapoport–Zink spaces} (see [RZ96]) and can be defined for all reductive linear algebraic groups over \( \mathbb{Q}_p \). He proposed to use them to construct \textit{local} Langlands parameters for all reductive linear algebraic groups over \( \mathbb{Q}_p \), along the lines of Vincent Lafforgue’s construction of global
Langlands parameters for global fields of positive characteristics (see [La18]), but in the much fancier category of diamonds (see [SW20] and [Sch17]).

Acknowledgements

Some materials in this article are based on introductory lectures given at the National Center for Theoretical Sciences (NCTS) in Taipei in 2014; at the Mathematical Sciences Research Institute (MSRI) in Berkeley in 2014; at the University of Minnesota, Twin Cities, in 2015; and at the summer school sponsored by ETH Zürich in Ascona, Switzerland, in 2016. We heartily thank the organizers for their invitations, and thank the audiences for their valuable feedbacks. We also thank Brian Conrad and Sug Woo Shin, and the anonymous referees, for their helpful comments and suggestions.

References


[Gor05] M. Goersky, Compactifications and cohomology of modular varieties, in Arthur et al. [AEK05], pp. 551–582.
[Hai05] T. J. Haines, Introduction to Shimura varieties with bad reductions of parahoric type, in Arthur et al. [AEK05], pp. 583–658.
[Har90] M. Harris, Automorphic forms and the cohomology of vector bundles on Shimura varieties, in Clozel and Milne [CM90], pp. 41–91.


AN EXAMPLE-BASED INTRODUCTION TO SHIMURA VARIETIES


[12b] ———, Geometric modular forms and the cohomology of torsion automorphic sheaves, in Ji et al. [JPYY12], pp. 183–208.


**Index**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>⟨·, ·⟩</td>
<td>5, 17, 31, 36, 51, 54</td>
</tr>
<tr>
<td>⟨·, ·⟩σ</td>
<td>31</td>
</tr>
<tr>
<td>0n</td>
<td>5</td>
</tr>
<tr>
<td>1a,b</td>
<td>6, 18, 23</td>
</tr>
<tr>
<td>1a</td>
<td>5, 6</td>
</tr>
<tr>
<td>(1) (Tate twist)</td>
<td>51</td>
</tr>
<tr>
<td>A</td>
<td>27</td>
</tr>
<tr>
<td>A0</td>
<td>51, 53</td>
</tr>
<tr>
<td>A0'</td>
<td>51, 52</td>
</tr>
<tr>
<td>An</td>
<td>?</td>
</tr>
<tr>
<td>A∞</td>
<td>52</td>
</tr>
<tr>
<td>A∞,p</td>
<td>57</td>
</tr>
<tr>
<td>abelian motive</td>
<td>57</td>
</tr>
<tr>
<td>abelian scheme</td>
<td>17, 50, 51, 53</td>
</tr>
<tr>
<td>universal</td>
<td>56</td>
</tr>
<tr>
<td>abelian variety</td>
<td>15, 17, 51, 56, 59</td>
</tr>
<tr>
<td>dual</td>
<td>52</td>
</tr>
<tr>
<td>abelian-type Shimura datum</td>
<td>53, 59, 62</td>
</tr>
<tr>
<td>Shimura variety</td>
<td>50, 62, 63</td>
</tr>
<tr>
<td>absolute Hodge cycle</td>
<td>59</td>
</tr>
<tr>
<td>adelic topology</td>
<td>4</td>
</tr>
<tr>
<td>adelic space</td>
<td>64</td>
</tr>
<tr>
<td>adjoint quotient</td>
<td>9, 18, 60, 61</td>
</tr>
<tr>
<td>adjoint representation</td>
<td>9, 52</td>
</tr>
<tr>
<td>adjoint Shimura datum</td>
<td>60, 61</td>
</tr>
<tr>
<td>affine group scheme</td>
<td>4</td>
</tr>
<tr>
<td>Albert algebra</td>
<td>27</td>
</tr>
<tr>
<td>Albert’s classification</td>
<td>55</td>
</tr>
<tr>
<td>algebraic curve</td>
<td>9, 34</td>
</tr>
<tr>
<td>algebraic cycle</td>
<td>50</td>
</tr>
<tr>
<td>algebraic group</td>
<td>4, 10, 10</td>
</tr>
<tr>
<td>simply-connected</td>
<td>25, 61</td>
</tr>
<tr>
<td>algebraic space</td>
<td>17, 47</td>
</tr>
<tr>
<td>algebraic stack</td>
<td>17, 50</td>
</tr>
<tr>
<td>algebraicity</td>
<td>9, 11, 17, 33, 34, 47, 49, 64</td>
</tr>
<tr>
<td>α0,ν</td>
<td>52, 53</td>
</tr>
<tr>
<td>αN</td>
<td>11</td>
</tr>
<tr>
<td>analytification</td>
<td>10, 11, 17, 17, 63</td>
</tr>
<tr>
<td>approximation</td>
<td>6</td>
</tr>
<tr>
<td>arithmetic subgroup</td>
<td>8, 10, 12, 37, 39</td>
</tr>
<tr>
<td>neat</td>
<td>8</td>
</tr>
<tr>
<td>Artin’s criterion</td>
<td>17</td>
</tr>
<tr>
<td>Aut(ℂ/F0),</td>
<td>55, 59</td>
</tr>
<tr>
<td>automatic compactness</td>
<td>34, 11, 42</td>
</tr>
<tr>
<td>automorphic L-function</td>
<td>58</td>
</tr>
<tr>
<td>automorphic line bundle</td>
<td>49</td>
</tr>
<tr>
<td>automorphic representation</td>
<td>3, 4, 48</td>
</tr>
<tr>
<td>automorphic vector bundle</td>
<td>49, 52</td>
</tr>
<tr>
<td>automorphy condition</td>
<td>49</td>
</tr>
<tr>
<td>automorphy factor</td>
<td>7, 49</td>
</tr>
</tbody>
</table>
INDEX 73

B, 11, 43, 54, 60
bad reduction, 57
blowup, 63
Borel subgroup, 35
Borel–Serre compactification, 47, 49
boundary, 34, 48, 50

codimension, 47, 49
stratification, 48, 68, 69

boundary component, 4, 34, 35, 37, 39,
41, 42, 44, 46, 47, 48, 49
bounded realization, 14, 18, 22, 25, 30

C, 64
canonical bundle, 49
canonical extension, 49, 57
canonical model, 3, 4, 11, 12, 49, 50
integral, 62
Cartan involution, 9
Cartan’s classification of Hermitian symmetric domain

type I, 21

type II, 22

type III, 33

type IV, 25

type V, 31

type VI, 31
Cayley numbers, 20, 26
Cayley transform, 15, 20, 30
Cent (centralizer), 16, 21, 26, 30, 32, 49

center of unipotent radical, 39, 41, 43
Chow’s theorem, 47
classical group, 35
Clifford algebra, 25
closed complex analytic subspace, 47
CM elliptic curve, 12
CM field, 41
CM point, 12, 55
cocenter, 55, 60
coherent cohomology, 50, 57
cohomology, 48, 50
coherent, 50, 57
étale, 11, 63, 84
intersection, 48
L², 48
nearby cycle, 57, 63
relative, 46
weighted, 49
commensurable, 6
compact factor, 9
compact Riemann surface, 9, 34
compaction
Borel–Serre, 47, 49
reductive Borel–Serre, 47, 49
Satake–Baily–Borel or minimal, 4, 11
integral model, 17, 56, 58, 63
toroidal, 47, 50
integral model, 56, 58, 63
complete algebraic curve, 9, 34
complex analytic space, 47
complex analytic subspace, 47
complex analytification, 10, 11, 14, 27
63
complex manifold, 7, 9
complex structure, 7, 17, 51
complex torus, 16, 51
cone decomposition, 47, 63
refinement, 48
congruence subgroup, 8, 9, 12, 17
principal, 8, 9, 11
conjugacy class, 9, 10, 12, 16, 21, 22
20, 30, 32, 33, 54, 59
connected, 6, 8, 10, 54, 55
connected component, 4, 8, 10, 12
23, 25, 34, 55, 60, 62
connected Shimura datum, 10, 10, 12
10, 12
connected Shimura variety, 12, 56, 63
collection, 50
cotangent bundle, 49
counting point, 57
cusp, 8, 34
cusp form, 10, 50
cyclic isogeny, 57
D, 6, 9, 54
D_a, b, 18, 21
D^±, 18, 21
D_a, b, 21
pu, 60
D_b, 31
D_b, 30, 31
D_a, 15
D^n, 8, 10
D^+(r), 44, 42
D_o, 22
D_o, 2, 25
D^+(o), 25
D^*, 34, 35, 47
de Rham, 64
degeneration, 56
Deligne torus, 9

<table>
<thead>
<tr>
<th>Concept</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deligne–Mumford stack</td>
<td>17, 56</td>
</tr>
<tr>
<td>derived group</td>
<td>28, 54, 55, 60, 61</td>
</tr>
<tr>
<td>determinantal condition</td>
<td>53</td>
</tr>
<tr>
<td>diamond</td>
<td>85</td>
</tr>
<tr>
<td>division algebra</td>
<td>12</td>
</tr>
<tr>
<td>normed</td>
<td>20</td>
</tr>
<tr>
<td>octonion</td>
<td>25</td>
</tr>
<tr>
<td>double coset space</td>
<td>17, 54, 55, 58</td>
</tr>
<tr>
<td>Drinfeld level structure</td>
<td>57, 58</td>
</tr>
<tr>
<td>dual abelian variety</td>
<td>52</td>
</tr>
<tr>
<td>dual lattice</td>
<td>52</td>
</tr>
<tr>
<td>Dynkin diagram</td>
<td>32, 33, 60</td>
</tr>
<tr>
<td>$E$</td>
<td>39, 54</td>
</tr>
<tr>
<td>$E_6$</td>
<td>30, 33</td>
</tr>
<tr>
<td>$E_6(-14)$</td>
<td>31–33</td>
</tr>
<tr>
<td>$E_6(-26)$</td>
<td>27, 32</td>
</tr>
<tr>
<td>$E_6(C)$</td>
<td>32</td>
</tr>
<tr>
<td>$E_7(-25)$</td>
<td>27, 32</td>
</tr>
<tr>
<td>$F$</td>
<td>41, 60</td>
</tr>
<tr>
<td>$F_0$</td>
<td>10, 49, 53, 54, 64</td>
</tr>
<tr>
<td>$f_4$</td>
<td>28</td>
</tr>
<tr>
<td>$F^+$</td>
<td>41</td>
</tr>
<tr>
<td>failure of Hasse’s principle</td>
<td>55, 56</td>
</tr>
<tr>
<td>faithful representation</td>
<td>4, 8, 58</td>
</tr>
<tr>
<td>fan</td>
<td>47</td>
</tr>
<tr>
<td>flag</td>
<td>30</td>
</tr>
<tr>
<td>flat</td>
<td>57, 58</td>
</tr>
<tr>
<td>Fontaine–Mazur conjecture</td>
<td>64</td>
</tr>
<tr>
<td>formally real Jordan algebra</td>
<td>27, 29</td>
</tr>
<tr>
<td>free action</td>
<td>5</td>
</tr>
<tr>
<td>Freudenthal’s theorem</td>
<td>30</td>
</tr>
<tr>
<td>fundamental domain</td>
<td>54</td>
</tr>
<tr>
<td>fundamental group</td>
<td>65</td>
</tr>
<tr>
<td>$G$</td>
<td>28, 30, 31, 33</td>
</tr>
<tr>
<td>$G_0$</td>
<td>10, 12</td>
</tr>
<tr>
<td>$G^\text{ad}(\mathbb{R})^+$</td>
<td>10</td>
</tr>
<tr>
<td>$(G, \mathcal{D})$</td>
<td>7, 9, 11, 54, 58</td>
</tr>
<tr>
<td>$(G, D^+)$</td>
<td>10, 12</td>
</tr>
<tr>
<td>$G^\text{der}$</td>
<td>50, 61</td>
</tr>
<tr>
<td>$G(\mathbb{Q})$</td>
<td>55</td>
</tr>
<tr>
<td>$G_m$</td>
<td>1</td>
</tr>
<tr>
<td>$G(\mathbb{Q})_\pm$</td>
<td>8</td>
</tr>
<tr>
<td>$C(r)$</td>
<td>44, 48</td>
</tr>
<tr>
<td>$G(R)^+$</td>
<td>28, 53, 58, 60</td>
</tr>
<tr>
<td>$G(R)^+$</td>
<td>8</td>
</tr>
<tr>
<td>$g^*$</td>
<td>28</td>
</tr>
<tr>
<td>$G_a$</td>
<td>50</td>
</tr>
<tr>
<td>GAGA</td>
<td>47</td>
</tr>
<tr>
<td>Gal($\mathbb{Q}/F_0$)</td>
<td>11</td>
</tr>
<tr>
<td>Galois cohomology</td>
<td>25</td>
</tr>
<tr>
<td>Galois representation</td>
<td>2, 11</td>
</tr>
<tr>
<td>geometric</td>
<td>64</td>
</tr>
<tr>
<td>Galois symmetry</td>
<td>8, 11</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>10, 11, 37, 59, 62, 64, 68</td>
</tr>
<tr>
<td>$\Gamma_0(p)$</td>
<td>57</td>
</tr>
<tr>
<td>$\Gamma\backslash D^+$</td>
<td>47, 49</td>
</tr>
<tr>
<td>$(\Gamma\backslash D)^{\text{cusp}}$</td>
<td>47, 49</td>
</tr>
<tr>
<td>$(\Gamma\backslash D)^{\text{min}}$</td>
<td>47, 49</td>
</tr>
<tr>
<td>$\Gamma\backslash D^+$</td>
<td>10, 11</td>
</tr>
<tr>
<td>$(\Gamma\backslash D)^{\text{cusp}}$</td>
<td>47, 49</td>
</tr>
<tr>
<td>$(\Gamma\backslash D)^{\text{min}}$</td>
<td>47, 49</td>
</tr>
<tr>
<td>$\Gamma\backslash D^*$</td>
<td>47, 49</td>
</tr>
<tr>
<td>$\Gamma_i$</td>
<td>8</td>
</tr>
<tr>
<td>$\Gamma(N)$</td>
<td>9, 11</td>
</tr>
<tr>
<td>general linear group</td>
<td>4, 40, 50, 58</td>
</tr>
<tr>
<td>general spin group</td>
<td>59, 61</td>
</tr>
<tr>
<td>generalized projective coordinates</td>
<td>13</td>
</tr>
<tr>
<td>geometric Galois representation</td>
<td>64</td>
</tr>
<tr>
<td>geometric invariant theory</td>
<td>17</td>
</tr>
<tr>
<td>GL$_1$</td>
<td>4</td>
</tr>
<tr>
<td>GL$_2$</td>
<td>3, 50, 58</td>
</tr>
<tr>
<td>GL$_{2a}(B)$</td>
<td>23</td>
</tr>
<tr>
<td>GL$_{2b}$</td>
<td>3</td>
</tr>
<tr>
<td>GL$_{r}(B)$</td>
<td>43</td>
</tr>
<tr>
<td>global Langlands correspondence</td>
<td>3</td>
</tr>
<tr>
<td>global Langlands parameter</td>
<td>65</td>
</tr>
<tr>
<td>good prime</td>
<td>53, 54, 56, 57</td>
</tr>
<tr>
<td>good reduction</td>
<td>53</td>
</tr>
<tr>
<td>Gram–Schmidt process</td>
<td>23</td>
</tr>
<tr>
<td>Grassmannian</td>
<td>14</td>
</tr>
<tr>
<td>group scheme</td>
<td>4</td>
</tr>
<tr>
<td>growth condition</td>
<td>49</td>
</tr>
<tr>
<td>GS$_{2n}$</td>
<td>7, 54, 55, 60, 61</td>
</tr>
<tr>
<td>GU$_{a,b}$</td>
<td>21, 74</td>
</tr>
<tr>
<td>GU$_{n,b}$</td>
<td>21, 74</td>
</tr>
<tr>
<td>$H$</td>
<td>7, 8, 13, 15, 25, 34, 48, 55</td>
</tr>
</tbody>
</table>
isomorphism class, 53
Iwasawa main conjecture, 3
J, 29
J2, 36
Ja,b, 19
Jn, 5
Jordan algebra, 27, 29, 30
K, 14, 18, 22, 26, 30, 32, 33
K3 surface, 61
Klingen parabolic subgroup, 36
Kottwitz conjecture, 58, 63
Kottwitz–Rapoport stratification, 58
Kuga family, 56
Kuga–Satake construction, 59, 61
Kuga–Sato variety, 50
L, 51
L2-cohomology, 48
L(I(Q), 44
L#, 51
L(r)(Q), 38
λ0, 51
Langlands parameter, 54
Langlands program, 3
Langlands–Rapoport conjecture, 58, 63
lattice, 10
level structure, 11
level, 14
Levi quotient, 44
Levi subgroup, 38
Hermitian part, 38
Lie A0, 52
Lie algebra condition, 53
light cone, 24
linear algebraic group, 4
local analogue of Shimura variety, 54
local Langlands conjecture, 5
local Langlands parameter, 54
local model, 57
local system, 6
étale, 63
M, 27
Ma,b, 18
Mn, 64
Mt, 55
manifold, 4
manifold, complex, 4
maximal compact subgroup, 16
maximal parabolic subgroup, 28
minimal compactification, 4
integral model, 17, 26, 58
mixed Hodge structure, 50
mixed Shimura variety, 50
integral model, 56
mixed type D, 61
mixed-characteristic shtuka, 64
Möbius transformation, 7
modular curve, 9
modular form, 9
p-adic, 57
modular function, 49
moduli interpretation, 58
moduli problem, 17
PEL, 51
morphism of Shimura data, 12
morphism of Shimura varieties, 12
motive, 64
multichain of isogenies, 57
multichain of lattices, 57
N(X), 26
nearby cycle, 57
neat, 8
Newton stratification, 68
normal, 37
normal crossings divisor, 48
normalization, 57
ν, 5
ν0,a, 52
number field, 9
O, 51
Oa,b, 6
Ok, 51
Of, 41
O(F0,1−(p), 64
O(F0,0−(p), 62
O-lattice, 51
On, 6
octonion division algebra, 24

octonion upper half-space, 29
Oort central leaves, 58
open compact subgroup, 7, 8, 10, 12
hyperspecial, 54, 62, 64
neat, 8, 12, 55
parahoric, 57, 62
orthogonal group, 9
orthogonal Shimura variety, 54, 55, 62

P, 28
$P_f(Q)$, 44
$p^-$, 9
$p^+$, 9
$P^{(r)}(Q)$, 38, 40, 42
$p$-adic $L$-function, 3
$p$-adic modular form, 57, 58
$p$-adic realization, 43
$p$-rank stratification, 58
parabolic subgroup, 35, 48
Klingen, 36
maximal, 35, 37, 40, 42, 44
minimal, 36
Siegel, 36
parahoric, 57, 62
parameter space, 11, 17, 51, 56, 59
partial minimal compactification, 58
partial toroidal compactification, 58
PEL datum
integral, 51, 53
rational, 63
PEL moduli problem, 51, 53, 55, 57
of type A, 55
of type C, 55
of type D, 55
PEL structure, 51, 53, 56
PEL-type
Shimura datum, 54, 69
Shimura variety, 54, 56
perfectoid Shimura variety, 64
Picard modular surface, 41
Poincaré upper half-plane, 26
polarization, 17, 51, 53
polarized abelian scheme, 17
polarized abelian variety, 16, 17, 59
polarized K3 surface, 61
positive involution, 41, 51, 53, 56
principal congruence subgroup, 8, 9, 11
principal level structure, 52
principal polarization, 17
projective coordinates, 7, 13, 15, 18, 19
23, 37, 39, 42, 49
projective spectrum, 56
projective variety, 11, 37, 39, 47, 48
proper base change theorem, 63
$\mathbb{Q}$, 44
$\mathbb{Q}_p$, 64
quasi-projective scheme, 55
quasi-projective variety, 4, 10, 17, 63, 64
quaternion algebra, 42, 60
quaternion upper half-space, 23
Rapoport–Zink space, 64
rational boundary component, 4, 34, 35
rational Hodge structure, 60
rational PEL datum, 53
real analytic topology, 6, 8, 10, 25
real quadratic extension, 35
real torus, 16, 61
realization
bounded, 15, 18, 22, 25, 30, 31
unbounded, 13, 18, 22, 24, 25, 29, 31
reduction
bad, 57
good, 58
reduction theory, 34, 37
reductive algebraic group, 4, 9, 64
reductive Borel–Serre compactification, 47, 49
reductive Lie group, 18
refinement of cone decompositions, 48
reflex field, 10, 19, 53, 54, 64
relative cohomology, 56
representation
automorphic, 3, 7, 48
Galois, 3, 11
generic, 63
resolution of singularity, 48
restricted product, 4
restriction of scalars, 9, 35, 41
$\rho$, 25, 30
$\rho_B$, 25, 29
Riemann surface, 9, 74
Riemannian symmetric space, 48
noncompact of type A I, 6
noncompact of type E IV, 28
Rosati condition, 52
$S$, 19, 39, 40
$S^*$, 4
$S_2$, 7
$S_3$, 7
$S_n$, 7
Satake topology, 35, 37, 39, 42, 46, 47
Satake–Baily–Borel compactification, 4, 11, 37, 39, 42, 46, 47, 49
integral model, 17, 56, 58, 63
semi-abelian scheme, 56
semisimple algebra, 35, 37, 39, 42, 46, 47
semisimple algebraic group, 4, 11
semisimple Lie group, 18
Shimura curve, 31, 60
Shimura datum, 3, 4, 9, 10, 11, 16, 22, 49, 54, 55
adjoint, 60, 61
connected, 10–12, 16
morphism, 12
of abelian type, 33, 59–62
of Hodge type, 58–62
of PEL type, 54, 59
simple, 61
Shimura reciprocity law, 12
Shimura variety, 3, 4, 9, 11, 49, 50, 58
CM point, 12, 56
connected, 12, 63
integral canonical model, 62
integral model, 4, 49, 51, 56, 62, 64
level, 11
local analogue, 64
morphism, 12
of abelian type, 60, 62, 64
of Hodge type, 58, 62, 64
of PEL type, 54, 56
perfectoid, 64
special, 12
special subvariety, 12
zeros-dimensional, 12
shtuka, 63
Siegel embedding, 22, 26, 30, 32, 33, 58
59, 62
Siegel modular variety, 17, 18, 49, 50, 58
Siegel moduli scheme, 17, 49, 51, 56, 64
Siegel parabolic subgroup, 39
Siegel upper half-space, 13
Σ, 17
σ, 31
signature, 6, 18, 24, 41, 44
similitude character, 31
simple algebra, 35
simple normal crossings divisor, 48
simple Shimura datum, 61
simply-connected algebraic group, 25
51
SL₂, 7, 9, 11, 15, 34, 48
SL₂ × SL₂, 94
SLₙ, 4, 6, 8, 9
smooth, 10, 48, 53, 62, 64
so₁₀, 31, 33
SO₂, 7, 15, 25
SO₂n, 22, 33, 42, 54, 55, 61
so₂₂n, 22, 61
so₂₂n₋₂₂, 61
SO₄, 23
so₆₂, 61
so₇, 61
SO₆₂(R)⁺, 25, 33, 44, 60
SO₆₂, 6, 23, 24, 44
SO₈, 6, 22, 23
SO(Q⁺, Q), 44
Sp₂, 15
Sp₂ₙ(2), 23
Sp₂ₙ, 6, 13, 32, 35, 51, 55
special linear group, 4, 6, 7, 9, 11, 15
54, 55, 56
special node, 92
special orthogonal group, 6, 6, 7, 15, 22
23, 25, 33, 42, 44, 54, 55, 60, 62
special point, 12
special subvariety, 12
special unitary group, 18
Spin₂₂, 33
Spin₂₉, 61
Spin₄₂, 25, 26, 33
spin group, 25, 33, 59, 61, 62
spinor norm, 25
×, 51, 54
stratification
boundary, 48, 58, 63
Ekedahl–Oort, 58
Kottwitz–Rapoport, 58
Newton, 58
p-rank, 58
SU₄₂, 18
subcanonical extension, 50, 57
sufficiently small, 8
Sym₄₂, 6, 13
symmetric space
Hermitian, 4, 6, 9, 12, 13, 23, 24, 32, 33
noncompact of type A III, 21
noncompact of type BD I, 28
noncompact of type C I, 18
noncompact of type D III, 22
noncompact of type E III, 31
noncompact of type E VII, 31
irreducible, 13, 29, 33, 38
Riemannian, 48
noncompact of type A I, 6
noncompact of type E IV, 28
symmetry
Galois, 3, 11
Hecke, 3, 7, 11
symplectic group, 5, 11, 12, 13, 15, 32, 33, 51, 54, 55, 58, 60, 61
symplectic Shimura variety, 54, 60, 61
tamely ramified, 62
Tate twist, 51
topology
adelic, 4
of $H^*$, 34
real analytic, 8, 10, 25
Satake, 35, 37, 39, 42, 46
Zariski, 6
toroidal compactification, 47–50
integral model, 56, 58, 63
torsion cohomology class, 3, 64
 torsion subgroup, 11, 52
torus, 12
totally imaginary extension, 41
totally isotropic subspace, 35–38, 40, 43, 44
totally real field, 35, 41, 60
trace condition, 53
transitivity, 6–8, 14, 21, 22, 25, 30
tube domain, 13, 20, 23–25, 29
type A, 18–22, 32, 39–42, 55, 58, 61
type B, 23–26, 33, 44–46, 61
type C, 13–18, 32, 35–39, 55, 58, 61
type D, 22–26, 33, 42–46, 55, 61
type E$_6$, 31–33, 46, 61, 63
 type E$_7$, 26–31, 33, 46, 61, 63
unipotent radical, 38, 40, 43, 44
unit disc, 15
unitary group, 3, 10, 11, 15, 18, 19, 22
unitary Shimura variety, 54, 55, 61
universal abelian scheme, 56
universal elliptic curve, 50
unramified, 53, 61
upper half-plane
Poincaré, 7, 12, 13, 15, 25
upper half-space
Hermitian, 20
octonion, 20
quaternion, 28
Siegel, 13
$V$, 23, 24
$V_0$, 22
$V_0^*$, 52
$V_C$, 23, 24
$V(\wedge^r)(\mathbb{Q})$, 38, 40, 43
variation of Hodge structures, 3, 9, 12
 variation of mixed Hodge structures, 50
vector bundle, 9, 49, 57
$W$, 49
$W^*$, 49
$W^{un}$, 49
$W^{un}(\mathbb{Q})$, 38, 40, 43
$W^{sub}$, 50
weak approximation, 6
weight, 19
weighted cohomology, 49
Weil pairing, 11, 62
$\hat{X}_{G_{\mathfrak{A}}}$, 50
$X_{\mathfrak{A}}$, 9, 11, 13, 18, 49
$X_{\mathfrak{A},F_{0}}$, 62
$X_{\mathfrak{A},\mathfrak{O}_{F_{0},(\mathfrak{v})}}$, 62
$X^{(i)}_{\mathfrak{A}}$, 18
$X_{\mathfrak{A},\mathfrak{O}_{F_{0},(\mathfrak{v})}}$, 62
$\hat{Z}$, 4
$\hat{Z}^*$, 57
$Z(M)$, 28
Zarhin’s trick, 62
Zariski topology, 6
zero-dimensional Shimura varieties, 12
$\zeta_N$, 11
Zucker’s conjecture, 48