Lectures on locally symmetric spaces and arithmetic groups

Lizhen Ji
Department of Mathematics
University of Michigan
Ann Arbor, MI 48109

May 21, 2004

Contents

1 Introduction

Locally symmetric spaces arise from many different areas such as differential geometry, topology, number theory, automorphic forms, representation theory, complex analysis, dynamical systems, algebraic geometry and string theory. The most important class consists of quotients of symmetric spaces by arithmetic groups, for example, the moduli space of elliptic curves is the quotient of the upper half plane $\mathbf{H}$ by $SL(2, \mathbb{Z})$. In these lectures, we introduce and study locally symmetric spaces, arithmetic groups, and reduction theory by emphasizing applications in various areas such as algebraic geometry, number theory, sphere packing, differential equations. We give important examples of arithmetic groups such as arithmetic Fuchsian groups, the Hilbert modular groups, the Bianchi group, the Picard modular groups. Besides the usual classical reduction theory developed by Borel & Harish-Chandra and refined by Borel, we also recall the precise reduction theory, and the reduction theory via polyhedral cones for linear symmetric spaces. Such a presentation gives a more global picture of the reduction theories.

The rest of these notes are organized as follows. In §2, we first motivate the definition of symmetric spaces and locally symmetric spaces, and show how they are related to Lie groups and discrete subgroups. Then we discuss various applications of locally symmetric spaces using the example of $SL(2,\mathbb{Z})\backslash\mathbf{H}$, where $\mathbf{H} = SL(2,\mathbb{R})/SO(2)$ is the Poincare upper plane. In §3, we recall the definition and basic properties of algebraic groups, in particular the important notion of ranks. In §4, we recall the definition of arithmetic subgroups and study several important examples. In §5, we give a compactness criterion for the quotient of symmetric spaces by arithmetic groups. In §6, we discuss the reduction theories. In §7, we describe applications of the reduction theories to metric properties of locally symmetric spaces. In §8, we briefly mention applications of the reduction theory to the spectral theory of locally symmetric spaces. It should be emphasized that we have left out many interesting topics related to locally symmetric spaces.

Acknowledgments: These notes are based on various lectures at the Center of Mathematical Sciences, Zhejiang University. I would like to thank the center for its hospitality. I would also like to thank I. Dolgachev for helpful comments and references and L.Saper for helpful comments on the precise reduction theory. This work is partially supported by an NSF grant.

2 Examples and applications of locally symmetric spaces

Before we give the formal definition of symmetric and locally symmetric spaces, we discuss several possible descriptions of abundance of symmetries of spaces.

Let M be a complete Riemannian manifold. Denote the isometry group of M by $Is(M)$. One way to say that M has a lot of symmetries is that for any pair of points $p, q \in M$, there exists an isometry $g \in Is(M)$ such that

$$gp = q.$$

Clearly $\mathbb{R}^n$ with the standard Euclidean metric has this property. But it turns out to be too weak to make M symmetric and instead describes the larger class of homogeneous manifolds. There are many homogeneous manifolds which are not symmetric.

In $\mathbb{R}^n$, there are more symmetries. In fact, for any two points $p, q \in \mathbb{R}^n$, and two unit vectors $u \in T_p\mathbb{R}^n$, $v \in T_q\mathbb{R}^n$, there exists $g \in Is(\mathbb{R}^n)$ such that

$$gp = q, \quad gu = v.$$

Therefore, another way is to impose this condition on M, i.e, $Is(M)$ acts transitively on the set of unit vectors (the unit sphere bundle) of M. It turns out that this is too restrictive and describes the class of symmetric spaces of rank 1 and the Euclidean spaces. The correct definition is to impose the condition only for the case $p = q$ and $u = -v$.

We start with the formal definition of locally symmetric spaces. Let M be a complete Riemannian manifold. For any point $x \in M$, there exists a (normal) neighborhood U such that

1. every point in U is connected to x via a unique geodesic,
2. there exists a star-shaped domain $V \subset T_xM$ containing the origin 0 and symmetric with respect to 0 such that the exponential map $\exp : V \to U$ is a diffeomorphism.

On such a neighborhood U, there is a geodesic symmetry s_x defined by reversing geodesics passing through x, i.e., for any geodesic $\gamma(t)$, $t \in \mathbb{R}$, with $\gamma(0) = x$,

$$s_x(\gamma(t)) = \gamma(-t),$$

when $\gamma(t) \in U$.

In terms of the exponential map, we have the commutative diagram:

$$\begin{array}{ccc} U & \xrightarrow{s_x} & U \\ \exp \big\uparrow & & \big\uparrow \exp \\ V & \xrightarrow{-Id} & V \end{array}.$$

From this commutative diagram, it is clear that s_x is a diffeomorphism of U. Since $s_x \neq Id$ and $s_x^2 = Id$, s_x is involutive and called the local geodesic symmetry at x.

Definition 2.1 (1) A complete Riemannian manifold M is called locally symmetric if for any $x \in M$, the (local) geodesic symmetry s_x is a local isometry. (2) The manifold M is called a symmetric space if it is locally symmetric and every local isometry s_x extends to a global isometry of M.

If M is symmetric, then for all values of t,

$$s_x(\gamma(t)) = \gamma(-t).$$

Clearly, symmetric spaces are also locally symmetric spaces. But when people talk about locally symmetric spaces, they usually refer to the special class of locally symmetric spaces of finite volume, due to various applications indicated below. In these notes, we will often follow this tradition.

It can be checked easily that $M = \mathbb{R}^n$, $\mathbb{R}^n/\mathbb{Z}^n$ and other quotients of $\mathbb{R}^n$, the spheres S^{n-1} in $\mathbb{R}^n$ are all symmetric spaces.

Some natural problems about locally symmetric spaces are:

1. Why are the locally symmetric spaces special, important?
2. How to construct them? Are there many of them?
3. How to understand them? How to study their geometric properties and analysis on them?
4. How to use them in other areas besides the original motivations?

The short answers to these questions are:

1. They are very special Riemannian manifolds. In fact, locally symmetric spaces can be defined as Riemannian manifolds such that the covariant derivative of the curvature is zero.

   Many interesting moduli spaces in algebraic geometry and number theory are given by locally symmetric spaces: for example, the moduli spaces of polarized abelian varieties [Mu, Theorem 4.7], and the moduli spaces of abelian varieties with certain endomorphisms groups (see [Hu]), the moduli spaces of polarized K-3 surfaces and the related Enriques surfaces (see [BPV] [Lo1] [Lo2]), some configuration spaces of points (see [Yos]), and the moduli space of quadratic forms, and lattices in $\mathbb{R}^n$. They are also important in string theory, since Calabi-Yau manifolds are 3-dimensional analogues of elliptic curves and K-3 surfaces, and the mirror maps are often given by modular forms (see [Yu] [Do] and the references there).

   The monodromy group of some differential equations with regular singularities also gives rise to interesting discrete subgroups and the related uniformizations are given by locally symmetric spaces (see [Ho1] [Ho2] [Yo1] [Yo2] [DM]).

   Arithmetic subgroups also naturally arise as the component group of the diffeomorphism group of simply connected manifolds of dimension greater than or equal to 6 (see [Su, Theorem 13.3]), and results from algebraic groups and arithmetic subgroups are needed to prove the result in [Su] that the diffeomorphism type of a compact smooth manifold is determined up to finitely many possibilities by some algebraic invariants.

   In many subjects such as geometry, geometric topology, and dynamical systems, people are interested in rigid or extremal objects. Locally symmetric spaces are crucial in such problems, for example, the Mostow rigidity [Mos], Margulis superrigidity [Mar], Zimmer's program on non-linear actions (a generalization of Margulis superrigidity) [Zi], the minimal entropy rigidity in dynamical systems [BCG], the Novikov conjectures and the Borel conjecture in topology [FRR], and the rigidity of complex manifolds [Mok].

2. They can be systematically constructed using Lie groups, algebraic groups and arithmetic groups, and can be used to study such groups.

3. Geometry of locally symmetric spaces are closely related to algebraic structures in Lie groups and algebraic groups. For locally symmetric spaces of finite volume, the reduction theory is crucial to problems both in geometry and analysis on locally symmetric spaces.

4. They are natural spaces for Lie groups and arithmetic groups to act on and hence give rise to natural representations of the Lie groups. These representations are very important in representation theory and number theory, in particular, the Langlands program.

In these lectures, we mainly concentrate on locally symmetric spaces and study Problems (2) and (3).

We start with a brief description of Problem (2) on how to construct locally symmetric spaces using group theory.

Locally symmetric and symmetric spaces are related by the following (see [Bo1]):

Proposition 2.2 *If M is a (complete) locally symmetric space, then its universal covering space $X = \widetilde{M}$ with the lifted Riemannian metric is (globally) symmetric.*

Let $\Gamma = \pi_1(M)$ be the fundamental group of M. Then Γ acts isometrically and properly on X, and

$$M = \Gamma \backslash X.$$

Hence, locally symmetric spaces are quotients of symmetric spaces.

Let $G = Is^o(X)$ be the identity component of the isometry group $Is(X)$ of X. The following is well-known (see [Ji1] for example).

Proposition 2.3 *If X is a symmetric space, then G is a Lie group and acts transitively on X.*

Fix a basepoint $x_0 \in X$ and denote the stabilizer of x_0 in G by K:

$$K = \{g \in G \mid gx_0 = x_0\}.$$

Then K is a compact subgroup of G and

$$G/K \cong X, \quad gK \mapsto gx_0,$$

i.e., X is a homogeneous space. It should be pointed out that not all homogeneous Riemannian manifolds are symmetric. For example, the Lie group $SL(2,\mathbb{R})$, or any noncompact semisimple Lie group G, with a left invariant Riemannian metric is not a symmetric space. The basic reason is that this metric is not right invariant.

The fundamental group Γ acts isometrically on X and is a discrete subgroup of G. Hence any locally symmetric space M is of the form

$$M = \Gamma\backslash G/K,$$

where G is a (connected) Lie group, Γ a discrete subgroup, and K is a compact subgroup of G. Therefore, each locally symmetric space determines a triple (G, K, Γ).

The basic point in constructing locally symmetric spaces is that we can reverse the above process. There are three basic types of symmetric spaces: compact, non-compact and flat types, corresponding to the sectional curvatures being non-negative, non-positive and zero. In these notes, we concentrate on symmetric spaces of non-compact type and their quotients.

If G is a connected noncompact semisimple Lie group with finite center, $K \subset G$ a maximal compact subgroup, and then endowed with a G-invariant metric, the homogeneous space $X = G/K$ is a symmetric space of noncompact type (see [Ji1]). Any discrete, torsion free discrete subgroup Γ of G acts isometrically and fixed-point free on X, and the quotient $\Gamma\backslash X$ is a locally symmetric space. Such discrete groups Γ are often constructed via algebraic groups and given by arithmetic groups (see §4 below).

More generally we can take Γ to be any discrete subgroup of G, not necessarily torsion free. Then $\Gamma\backslash X$ is not necessarily smooth, but rather has finite quotient singularities, called V-manifolds or orbifolds. Since many natural important arithmetic groups such as $SL(2, \mathbb{Z})$ are not torsion free, we also call $\Gamma\backslash X$ a locally symmetric space for any non-torsion free discrete subgroup Γ.

In the rest of this section, we consider a simple example of locally symmetric spaces and answer the four questions raised at the beginning in more detail.

Consider

$$G = SL(2, \mathbb{R}) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{R}, ad - bc = 1\},$$

$$K = SO(2) = \{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \mid \theta \in \mathbb{R}\} \cong S^1,$$

$$\Gamma = SL(2, \mathbb{Z}) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathbb{Z}, ad - bc = 1\}.$$

The modular group $\Gamma = SL(2, \mathbb{Z})$ is not torsion free. In fact,

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \neq Id, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^4 = Id.$$

To get torsion free subgroups, for every $n \geq 1$, define the principal congruence subgroup

$$\Gamma_n = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}) \mid \begin{pmatrix} a & b \\ c & d \end{pmatrix} = Id \mod n\}.$$

When $n \geq 3$, Γ_n is torsion free. Though these congruence subgroups are important, we will concentrate on the modular group $SL(2, \mathbb{Z})$.

The first problem is to get a concrete realization of $X = G/K = SL(2, \mathbb{R})/SO(2)$. Let

$$\mathbf{H} = \{z = x + iy \in \mathbb{C} \mid x \in \mathbb{R}, y > 0\}$$

with metric $ds^2 = \frac{1}{y^2}(dx^2 + dy^2)$. Then $\mathbf{H}$ is a simply connected surface of constant curvature -1, the hyperbolic plane, i.e., a space form of constant curvature -1 in dimension 2. The group $SL(2, \mathbb{R})$ acts isometrically and holomorphically on $\mathbf{H}$ via fractional linear transformation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d} .$$

Using this transitive isometric action of $SL(2, \mathbb{R})$ on $\mathbf{H}$, we can show easily that $\mathbf{H}$ is a symmetric space. In fact, the geodesic symmetry at i is given by

$$s_i(z) = -1/\bar{z} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \cdot z$$

and is an isometry of $\mathbf{H}$. Under the conjugation by elements in $SL(2, \mathbb{R})$, it implies that for any point $x \in \mathbf{H}$, the geodesic symmetry i_x is a global isometry of $\mathbf{H}$ as well.

We remark that

$$Is^0(\mathbf{H}) = PSL(2, \mathbb{R}) = SL(2, \mathbb{R})/ \pm Id,$$

since $-Id$ acts trivially on $\mathbf{H}$. The stabilizer of the basepoint $x_0 = i$ in $SL(2, \mathbb{R})$ is equal to $K = SO(2)$, and hence

$$X = SL(2, \mathbb{R})/SO(2) \cong \mathbf{H}, \quad gSO(2) \mapsto gi$$

and the locally symmetric space associated with the triple (G, K, Γ) is $\Gamma\backslash\mathbf{H}$.

The next problem is to understand the structure of the quotient $\Gamma\backslash\mathbf{H}$. For this purpose, we need to introduce the notion of fundamental domains.

Definition 2.4 A fundamental domain for a discrete group Γ acting on $\mathbf{H}$ is an open subset set $\Omega \subset \mathbf{H}$ such that

1. Each coset $\Gamma \cdot x$ contains at least one point in the closure $\overline{\Omega}$, i.e., $\mathbf{H} = \Gamma\overline{\Omega}$.
2. No two interior points of Ω lie in one Γ-orbit, i.e., $\gamma\Omega$, $\gamma \in \Gamma$, are disjoint open subsets.

In some books, they also require Ω to be connected, but we do not impose this condition in these notes in view of the fact below that we often take unions of Siegel sets in constructing fundamental domains (sets).

Given a fundamental domain Ω, we can find a subset F, $\Omega \subset F \subset \overline{\Omega}$ such that each Γ-orbit contains exactly one point in F, and hence

$$\Gamma\backslash\mathbf{H} \cong F$$

as sets. Such a set F is called an *exact fundamental set* for Γ and is usually not open. If Ω is chosen nicely and the identification $\sim$ of the boundary $\partial\overline{\Omega}$ is known, we can understand the topology of $\Gamma\backslash\mathbf{H}$ as well by the homeomorphism

$$\Gamma\backslash\mathbf{H} \cong \overline{\Omega}/\sim,$$

where $\overline{\Omega}/\sim$ is given the quotient topology.

The next problem is to find a good fundamental domain for $SL(2,\mathbb{Z})$. This is given by reduction theory. For $SL(2,\mathbb{Z})$, it is well-known and simple. But for general X and Γ, it is complicated and one of the main topics of these lectures. As will be seen below, the reason why it is called reduction theory is that it is directly related to reduction of quadratic forms, motivated by problems of representations of integers by quadratic forms.

Proposition 2.5 *A fundamental domain Ω for $\Gamma = SL(2,\mathbb{Z})$ on $\mathbf{H}$ is given by the following region*

$$\Omega = \{z = x + iy \in \mathbf{H} \mid |z| > 1, -\frac{1}{2} < x < \frac{1}{2}\}.$$

Proof. There are two steps: (1) Using some extremal property to show that every Γ-orbit contains at least one point in $\overline{\Omega}$. (2) Show that no two points of Ω lie in one Γ-orbit.

Since Γ contains $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, which induces translation $z \to z+1$, it is clear that every Γ-orbit contains a point z with $\mathrm{Re}(z) \in [-\frac{1}{2}, \frac{1}{2}]$. For such a point z, consider the orbit Γz. To control the imaginary part, we choose $\gamma \in \Gamma$ such that $\mathrm{Im}(\gamma z)$ is maximal. If

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{then} \quad \mathrm{Im}(\gamma z) = \frac{y}{|cz+d|^2}.$$

Since a, b, c, d are integers, $c + dz$ is contained in the lattice $\mathbb{Z} + \mathbb{Z}z$ in $\mathbb{C} = \mathbb{R}^2$, and hence $\mathrm{Im}(\gamma z)$ is uniformly bounded and the maximum value is achieved. Since translation by $z \to z+1$ does not change the imaginary part, we can assume that for the point γz with the maximal imaginary, $\mathrm{Re}(\gamma z) \in [-\frac{1}{2}, \frac{1}{2}]$ and hence take $z = \gamma z$. Then for all $\gamma \in \Gamma$,

$$\mathrm{Im}(z) \geq \mathrm{Im}(\gamma z). \tag{1}$$

We claim that $|z| \geq 1$. Otherwise, $|z| < 1$. Take $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \Gamma$. Then

$$Sz = -\frac{1}{z} = -\frac{x - iy}{|z|^2} = -\frac{x}{|z|^2} + \frac{y}{|z|^2},$$

and hence

$$\mathrm{Im}(Sz) = \frac{y}{|z|^2} > \mathrm{Im}(z).$$

This contradicts Equation (1) and completes Step (1).

To prove Step (2), suppose that $z, \gamma z \in \Omega$ for some $\gamma \in \Gamma$. Assume that $\mathrm{Im}(\gamma z) \geq \mathrm{Im}(z)$. Then the above computations show that

$$\frac{y}{|cz+d|^2} \geq y,$$

and hence

$$|cz+d| \leq 1. \tag{2}$$

Since $|cz+d| \geq |c|\mathrm{Im}(z)$ and $\mathrm{Im}(z) \geq \frac{\sqrt{3}}{2} > \frac{1}{2}$, it implies that $|c| < 2$, i.e., $c = 0, \pm 1$. We now discuss various choices of c case by case.

(I). If $c = 0$, Equation (2) implies that $|d| \leq 1$. Then $ad - bc = 1$ implies that $ad = 1$ and $|d| = 1$. When $d = 1$, then $a = 1$, and $\gamma z = z + b$. Since $z, \gamma z \in \Omega$, $b = 0$, i.e., $\gamma = Id$. When $d = -1$, then $a = -1$, $\gamma z = -z - b$, and hence $b = 0$, $\gamma = -Id$.

(II) If $c = 1$, Equation (2) implies $|z+d| \leq 1$. Since $|\mathrm{Re}(z)| < \frac{1}{2}$ and d is integral, $|d| \leq 1$, and hence $d = 0, \pm 1$. The case $d = 0$ can not happen, otherwise $|z+d| = |z| > 1$ by the definition of Ω and contradicts Equation (2). The case $d = 1$ can not happen either, otherwise $|z+d| = |z+1| > |z| > 1$, since $|\mathrm{Re}(z+1)| > |\mathrm{Re}(z)|$. The case $d = -1$ is similarly excluded.

(III) The case $c = -1$ can be excluded by the same arguments or apply (II) to $-\gamma$. Hence $\gamma = \pm Id$. This completes Step (2).

There are several immediate applications of the identification of the fundamental domain Ω.

Corollary 2.6 *The group $SL(2, \mathbb{Z})$ is generated by*

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Proof. Note that $\mathbf{H}$ is covered by translates $\gamma\overline{\Omega}$, $\gamma \in \Gamma$, of $\overline{\Omega}$ which are reflections with respect to the sides of such domains. The domains which share common sides with $\overline{\Omega}$ are $T^{-1}(\overline{\Omega})$, $T(\overline{\Omega})$ and $S(\overline{\Omega})$. Hence any such domain is of the form $f(S,T)\overline{\Omega}$, where $f(S,T)$ is an element in the subgroup $< S, T >$ generated by S, T. It follows that for every point $z \in \mathbf{H}$, there exists $f(S,T)$ such that $f(S,T)z \in \overline{\Omega}$. Take $z_0 \in \Omega$. For any $\gamma \in \Gamma$, there exists an element $f(S,T)$ in $< S, T >$ such that $f(S,T)\gamma z_0 \in \overline{\Omega}$. Since $z_0 \in \Omega$,

$$f(S,T)\gamma z_0 = z_0.$$

The proof of the above proposition shows that $f(S,T)\gamma = \pm Id$. Since

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = S^2,$$

γ also belongs to $< S, T >$. Hence $\Gamma =< S, T >$, i.e., Γ is generated by S, T.

Remark 2.7 The fundamental domain Ω can be regarded as the Dirichlet domain for $\Gamma = SL(2,\mathbb{Z})$ with center iy_0, $y_0 > 1$, i.e.,

$$\Omega = \{z \in \mathbf{H} \mid d(z, iy_0) < d(\gamma z, iy_0), \gamma \in \Gamma\}.$$

For each γ, the bi-sector $\{z \in \mathbf{H} \mid d(z, iy_0) = d(\gamma z, iy_0)\}$ is a geodesic. The three sides of Ω are contained in the bi-sectors of T^{-1}, T, S.

Corollary 2.8 *The quotient* $\Gamma\backslash\mathbf{H}$ *is noncompact but has finite area.*

Proof. The fundamental domain Ω is contained in the subset

$$\mathcal{S} = \{z = x + iy \in \mathbf{H} \mid y > \frac{\sqrt{3}}{2}, -\frac{1}{2} < x < \frac{1}{2}\}.$$

Then

$$\mathrm{Area}(\Gamma\backslash\mathbf{H}) = \int_\Omega \frac{dxdy}{y^2} \leq \int_{\mathcal{S}} \frac{dxdy}{y^2} = \int_{\frac{\sqrt{3}}{2}}^{+\infty} \frac{dy}{y^2} < +\infty.$$

To show $\Gamma\backslash\mathbf{H}$ is noncompact, let t_j be a sequence of real numbers going to $+\infty$. Identify it_j with its image in $\Gamma\backslash\mathbf{H}$. Clearly it can not converge to any point in $\Gamma\backslash\mathbf{H}$.

Remark 2.9 The set $\mathcal{S}$ is called a Siegel set for Γ associated with the cusp $i\infty$ and plays an important role in the general reduction theory. It is slightly larger than a fundamental set, but the fibers of the map $\overline{\mathcal{S}} \to \Gamma\backslash\mathbf{H}$ are finite and uniformly bounded (in fact by 4 in this case). In the general case, two immediate applications of existence of a nice fundamental domain for arithmetic groups are similar to those above: (1) the finite generation of the arithmetic groups and (2) finite volume of the quotients by arithmetic groups if the algebraic groups are semisimple.

For the application to quadratic forms, we need an exact fundamental set.

Proposition 2.10 *Let F be the union of Ω and $\partial\Omega \cap \{z \in \mathbf{H} \mid Re(z) \leq 0\}$, i.e., the left half of the boundary $\partial\Omega$. Then F is an exact fundamental set of Γ, i.e., it intersects each Γ-orbit at one point.*

Proof. Since the only possible identifications among the boundary points of Ω are given by the translation $z \to z + 1$, and the inversion $z \to z^{-1}$ when $|z| = 1$, it is clear that the left side of $\partial\Omega$ is identified with the right half side.

Reduction of quadratic forms.

Next we explain relations between the determination of the fundamental domain Ω and the theory of reduction of binary quadratic forms, which was studied by Lagrange, Legendre, Gauss and others.

Let $f(u, v) = au^2 + buv + cv^2$, $a, b, c \in \mathbb{Z}$, be an integral binary quadratic form. There are two basic problems in number theory:

1. Find integers n that can be represented by f, i.e., $n = f(u,v)$ for some $u, v \in \mathbb{Z}$.

2. If n can be represented by f, determine the number of solutions of $n = f(u,v)$, $u, v \in \mathbb{Z}$, i.e., the number of ways to represent n by $f(u,v)$. This is also called the multiplicity of the representation of n.

Two quadratic forms $f(u,v)$, $g(u,v)$ are called *equivalent* if there exists $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in GL(2,\mathbb{Z})$ such that

$$g(u,v) = f(\alpha u + \beta v, \gamma u + \delta v), \tag{3}$$

i.e., under the linear transform

$$\begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \alpha u + \beta v \\ \gamma u + \delta v \end{pmatrix}, \tag{4}$$

f is mapped to g.

Proposition 2.11 *Two equivalent quadratic forms represent the same set of integers with the same multiplicity.*

Proof. Since the linear transformation in Equation (4) preserves integral vectors and has an inverse given by an integral matrix, the proposition is clear.

To remove the redundancy in each equivalence class, we need to pick a good representative. For this purpose, we define that two quadratic forms f, g are *properly equivalent* if the matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ in Equation (4) belongs to $SL(2,\mathbb{Z})$.

Hence, $SL(2,\mathbb{Z})$ acts on the space of quadratic forms, and each proper equivalence class is a $SL(2,\mathbb{Z})$-orbit. The problem of finding good representatives of the proper equivalence classes is equivalent to finding a set of representatives which is mapped bijectively to the quotient under $SL(2,\mathbb{Z})$. We will use the fundamental domain Ω of $SL(2,\mathbb{Z})$ acting on $\mathbf{H}$ to solve this problem.

To be explicit, the quadratic form $f(u,v) = au^2 + buv + cv^2$ corresponds to the symmetric matrix $\begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}$. Then the action of $SL(2,\mathbb{Z})$ on the space of quadratic forms corresponds to the following action:

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \cdot \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}. \tag{5}$$

Since the assumption that a, b, c are integral is not essential in the following discussions, we will allow them to be any real numbers and establish a correspondence between the space of quadratic forms and $\mathbf{H}$, which is equivariant with respect to the action of $SL(2,\mathbb{Z})$.

It is known that the quadratic form $f(u,v) = au^2 + buv + cv^2$ is positive definite if $a > 0$ and $d = b^2 - 4ac < 0$, where d is the discriminant of f. In the following we will only deal with positive definite quadratic forms. The quadratic equation

$$az^2 + bz + c = 0, \quad \text{where} \quad z = \frac{u}{v},$$

has two distinct complex roots

$$z = \frac{-b + i\sqrt{|d|}}{2a}, \quad \bar{z} = \frac{-b - i\sqrt{|d|}}{2a}.$$

The root $z \in \mathbf{H}$ and determines the form f up to a positive constant. In fact,

$$f(u, v) = a(u - zv)(u - \bar{z}v),$$

and the coefficient a is uniquely determined by the discriminant d.

Proposition 2.12 *For each $d < 0$, denote by $\mathcal{Q}_d$ the set of positive definite quadratic forms $f(u, v)$ with discriminant d. Then $\mathcal{Q}_d$ corresponds bijectively to $\mathbf{H}$ under the map*

$$f = au^2 + buv + cv^2 \mapsto z = \frac{-b + i\sqrt{|d|}}{2a},$$

and this map is equivariant with respect to the action of $SL(2, \mathbb{Z})$.

Proof. As explained above, this map is injective. Since for any $z \in \mathbf{H}$, $u - zv \neq 0$ for all $u, v \in \mathbb{R}$, $(u - zv)(u - \bar{z}v) = |u - zv|^2 > 0$, hence there exists a unique positive constant a such that $f(u, v) = a(u - zv)(u - \bar{z}v)$ is a positive definite quadratic form of discriminant d. This shows that this map is bijective. Since $z = u/v$ and the action of $SL(2, \mathbb{Z})$ on the quadratic forms is given by Equation (4), it is clear that the map is $SL(2, \mathbb{Z})$-equivariant.

Remark 2.13 In fact, $SL(2, \mathbb{R})$ acts transitively on the space of positive definite quadratic forms of a fixed discriminant and the above map is equivariant with respect to $SL(2, \mathbb{R})$.

Proposition 2.14 *Under the map $\mathcal{Q}_d \to \mathbf{H}$ in the previous proposition, the quadratic forms $f(u, v)$ corresponding to the exact fundamental domain $F \subset \overline{\Omega}$ in Proposition (2.10) satisfy*

1. *either $0 \leq b \leq a = c$,*
2. *or $-a < b \leq a < c$.*

Proof. By definition, $z = \frac{-b+i\sqrt{4ac-b^2}}{2a}$, and $\mathrm{Re}(z) = -\frac{b}{2a}$. If $z \in F$, then

$$-\frac{1}{2} \leq -\frac{b}{2a} < \frac{1}{2},$$

hence

$$-a < b \leq a.$$

Since

$$|z|^2 = \frac{b^2 - d^2}{4a^2} = \frac{4ac}{4a^2} = \frac{c}{a},$$

$|z| \geq 1$ implies that $c \geq a$, and $|z| > 1$ implies that $c > a$. (Recall that the quadratic forms are positive and hence $a > 0$.) If $c = a$, then $|z| = 1$, and the condition $-\frac{1}{2} \leq \mathrm{Re}(z) \leq 0$ is equivalent to

$$0 \leq b \leq a.$$

These conditions are exactly the conditions given in the proposition.

Quadratic forms $f(u, v) = au^2 + buv + cv^2$ whose coefficients a, b, c satisfying the conditions in the above proposition are called reduced forms in number theory (see [Cox] [Gol]). Each proper equivalence class of positive definite quadratic forms contains exactly one reduced form.

Remark 2.15 For each fixed discriminant d, there are only finitely many reduced positive definite integral quadratic forms, and they can be listed. This is important for the problem of representations of integers. Given any integer n, it is easy to decide whether there exists a form of discriminant d that represents n. In fact, the precise condition is $d = b^2 \mod 4|n|$ for some b, i.e., d is a square residue mod $4|n|$. These two steps allow one to decide whether a given quadratic form represents an integer.

Geometry of numbers and sphere packing.

A lattice Λ in $\mathbb{R}^n$ is a discrete subgroup of rank n,

$$\Lambda = \mathbb{Z}v_1 + \cdots + \mathbb{Z}v_n,$$

where $v_1, \cdots, v_n$ are linearly independent vectors in $\mathbb{R}^n$. (The lattices defined here correspond to full lattices in some books.)

A fundamental domain for Λ acting on $\mathbb{R}^n$ is

$$(0, 1)v_1 + \cdots + (0, 1)v_n,$$

and hence

$$vol(\mathbb{R}^n/\Lambda) = |\det A|, \quad A = (v_1, \cdots, v_n).$$

For each lattice Λ, there is a sphere packing by placing a sphere of a common radius r at each lattice point so that these spheres do not overlap and r is chosen to be maximum with respect to this non-overlapping property. Since there is one sphere for each vertex of Λ, the density of the sphere packing is equal to

$$\frac{\sigma_n r^n}{vol(\mathbb{R}^n/\Lambda)},$$

where $\sigma_n r^n$ is the volume of the ball of radius r in $\mathbb{R}^n$. Clearly, this density is invariant under scaling.

A basic problem in the theory of sphere packing is to find a lattice Λ with the maximum density. We can find the best lattice packing in the case $n = 2$ by using the fundamental domain Ω for $SL(2, \mathbb{Z})$.

The first problem is to determine r for each lattice Λ. Let

$$a = \min_{v \in \Lambda, v \neq 0} ||v||^2,$$

the minimum norm square of nonzero vectors in Λ, then

$$r = \frac{\sqrt{a}}{2}.$$

The problem of densest packing becomes a problem to find the (global) maximum of the ratio

$$\frac{a^{n/2}}{vol(\mathbb{R}^n/\Lambda)}$$

among all lattices.

To solve this problem, we need to parametrize the set of lattices in $\mathbb{R}^n$. Each basis $v_1, \cdots, v_n$ of $\mathbb{R}^n$ determines a lattice $\Lambda = \mathbb{Z}v_1 + \cdots + \mathbb{Z}v_n$. But different bases can give rise to the same lattice. In fact, two bases $v_1, \cdots, v_n$ and $w_1, \cdots, w_n$ generate the same lattice if and only if there exists an element $T \in GL(n, \mathbb{Z})$ such that

$$(w_1, \cdots, w_n) = T(v_1, \cdots, v_n).$$

Two such bases are called equivalent bases. To parametrize lattices in terms of bases, we need to choose "good" (or reduced) bases.

In fact, such reduced bases are related to reduced quadratic forms. For each lattice $\Lambda = \mathbb{Z}v_1 + \cdots + \mathbb{Z}v_n$ corresponding to the basis $v_1, \cdots, v_n$, let $A = (v_1, \cdots, v_n)$ be the corresponding $n \times n$ matrix. Define a symmetric matrix

$$S = A\ {}^tA,$$

where ${}^tA$ is the transpose of A. Then S determines a positive definite quadratic form

$$Q(X, Y) = {}^tXSY, \quad X, Y \in \mathbb{R}^n.$$

On the other hand, the quadratic form Q determines the matrix A and the lattice Λ up to rotation on the right. Since the density of the lattice packing is invariant under rotation, we can ignore this non-uniqueness. The minimum norm square of nonzero vectors in Λ is equal to the minimum value of Q on the non-zero integral vectors. Let $<, >$ be the standard Euclidean quadratic form on $\mathbb{R}^n$, and $\mathbb{Z}^n$ the standard lattice. Let $\mathcal{P}_1$ be the space of pairs $(<, >, \Lambda)$ of the standard quadratic form and arbitrary lattices, and $\mathcal{P}_2$ the space of pairs $(Q, \mathbb{Z}^n)$ of arbitrary quadratic forms and the standard lattice $\mathbb{Z}^n$. Then the above transform can be viewed as a correspondence between these spaces $\mathcal{P}_1, \mathcal{P}_2$ of pairs of quadratic forms and lattices.

The above correspondence shows that the problem of maximum density becomes a problem to find a positive definite quadratic form of determinant 1 whose minimum value on $\mathbb{Z}^n$ is maximal. In dimension $n = 2$, this problem can be solved by using reduced forms.

For a reduced binary quadratic form $au^2 + buv + cv^2$, the coefficients a, b, c satisfy:

$$0 < a \leq c, \quad -a < b \leq a.$$

Hence

$$a = \min_{(u,v)\neq 0, u,v\in\mathbb{Z}} f(u, v).$$

Let ρ be the density of the lattice packing corresponding to f. Then

$$\rho = \frac{\pi r^2}{\sqrt{ac - \frac{b^2}{4}}} = \frac{\pi \frac{a}{4}}{\sqrt{ac - \frac{b^2}{4}}},$$

$$\rho^2 = \frac{\pi^2 a^2}{4(4ac - b^2)} \leq \frac{\pi^2 a^2}{4(4a^2 - a^2)} = \frac{\pi^2}{12}.$$

The maximum density

$$\rho = \frac{\pi}{2\sqrt{3}}$$

is achieved when $a = b = c$.

We need to determine the corresponding lattice $\Lambda = \mathbb{Z}v_1 + \mathbb{Z}v_2$. Since $A = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$, $S = A\,{}^tA$, and $S = \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}$, it follows that

$$a = |v_1|^2, \quad c = |v_2|^2, \quad b = 2 < v_1, v_2 > .$$

Let θ be the angle between v_1 and v_2. Then

$$\cos\theta = \frac{< v_1, v_2 >}{||v_1||\,||v_2||} = \frac{b/2}{\sqrt{ac}} \leq \frac{1}{2},$$

and the upper bound is achieved when $a = b = c$, which corresponds to the densest sphere packing as concluded above. Therefore, the lattice Λ having the maximal density is equal to $\mathbb{Z}v_1 + \mathbb{Z}v_2$ up to scaling and rotation, where

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix}.$$

Intuitively, this densest sphere packing is obtained by placing the next row of spheres in the holes of the previous row, which is apparently denser than the lattice packing of the standard lattice $\mathbb{Z}^n$.

Moduli space of elliptic curves.

We have identified the locally symmetric space $\Gamma\backslash\mathbf{H}$ as the set of equivalence classes of positive definite binary quadratic forms of determinant one, and also as the space of lattices of co-volume 1 up to rotation. Next we will show that it is also the moduli space of complex elliptic curves.

Definition 2.16 *An (complex) elliptic curve is a compact smooth Riemann surface Σ (or algebraic curve over $\mathbb{C}$) of genus 1.*

By the Riemann uniformization theorem, an elliptic curve Σ is of the form $\mathbb{C}/\Lambda$, where Λ is a lattice in $\mathbb{C}$. Two elliptic curves Σ_1, Σ_2 are called *equivalent* if there exists a biholomorphic map $\varphi : \Sigma_1 \to \Sigma_2$. The moduli space of elliptic curves is the set of equivalence classes of elliptic curves.

Proposition 2.17 *The moduli space of elliptic curves can be identified with $SL(2,\mathbb{Z})\backslash\mathbf{H}$ by the map $z \in \mathbf{H} \mapsto \mathbb{C}/\mathbb{Z} + \mathbb{Z}z$.*

To prove this proposition, we need the following lemma.

Lemma 2.18 *Two elliptic curves $\mathbb{C}/\Lambda_1, \mathbb{C}/\Lambda_2$ are equivalent if and only if there exists a nonzero constant a such that $a\Lambda_1 = \Lambda_2$.*

Proof. Let $\varphi : \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2$ be a biholomorphic map. Then it lifts to a holomorphic map $\tilde{\varphi} : \mathbb{C} \to \mathbb{C}$ with $\varphi(0) = 0$. Similarly, the inverse map $\varphi^{-1} : \mathbb{C}/\Lambda_2 \to \mathbb{C}/\Lambda_1$ lifts to a holomorphic map $\widetilde{\varphi^{-1}} : \mathbb{C} \to \mathbb{C}$ with $\widetilde{\varphi^{-1}}(0) = 0$, and is the inverse of $\widetilde{\varphi}$. Therefore, $\widetilde{\varphi}$ is a biholomorphic map of $\mathbb{C}$. We claim that $\widetilde{\varphi}$ must be linear, $\widetilde{\varphi}(z) = az + b$, and the condition $\widetilde{\varphi}(0) = 0$ implies that $\widetilde{\varphi}(z) = az$ and proves the proposition.

To prove the claim, consider the analytic power series

$$\widetilde{\varphi}(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Since $\widetilde{\varphi}$ is univalent, by the small Picard theorem, $\widetilde{\varphi}$ does not have essential singularities, and hence $a_n = 0$ for $n \gg 0$, and hence $\widetilde{\varphi}$ is a polynomial. By the univalence again, the degree of the polynomial is equal to 1. Therefore, the claim is proved.

Now we prove the proposition. Any lattice Λ in $\mathbb{C}$ is of the form $\mathbb{Z}v_1 + \mathbb{Z}v_2$, where $v_1, v_2 \in \mathbb{C}$ are linearly independent over $\mathbb{R}$, i.e., v_1/v_2 is not real. Since $(v_1/v_2)(v_2/v_1) = 1$, one of them, say v_1/v_2, has positive imaginary part and hence belongs to $\mathbf{H}$. Let $z = v_1/v_2$. Since the bases of a lattice are acted upon by $SL(2,\mathbb{Z})$, the proposition follows.

Monodromy groups of hypergeometric differential equations

The modular group $SL(2,\mathbb{Z})$ and its congruence subgroups also arise naturally in the study of hypergeometric differential equations. In fact, the familiar concept of Fuchsian groups in the theory of Riemann surfaces arose from the monodromy group of ordinary differential equations with regular singularities and the related uniformization. The references for the discussions here are [Ho1] [Ho2] [Yo1] [Yo2].

Recall that the hypergeometric differential equations are given by

$$z(1-z)w'' + (c - (a+b+1)z)w' - abw = 0,$$

where $z \in \mathbb{C}P^1$, a, b, c are constants, and $w = w(z)$ is a meromorphic function of z. This differential equation has regular singularities at three points $z = 0, 1, \infty$.

Fix any basepoint $z_0 \neq 0, 1, \infty$, and any two linearly independent solutions $w_1 = w_1(z)$, $w_2 = w_2(z)$ in a neighborhood of z_0. Then analytic continuation of w_1, w_2 along paths in $\mathbb{C}P^1 - \{0, 1, \infty\}$ defines multivalued functions on $\mathbb{C}P^1 - \{0, 1, \infty\}$ and gives a monodromy representation

$$\rho : \pi_1(\mathbb{C}P^1 - \{0, 1, \infty\}) = \pi_1(\mathbb{C}P^1 - \{0, 1, \infty\}, z_0) \to GL(2, \mathbb{C}).$$

This representation depends on the choice of the basepoint point z_0 and the solutions w_1, w_2. But different choices lead to conjugate representations. Hence, the differential equation determines a unique conjugacy class of representation of $\pi_1(\mathbb{C}P^1 - \{0, 1, \infty\})$ in $GL(2, \mathbb{C})$. We will also denote the representation into $PGL(2, \mathbb{C})$ by ρ. The image $\rho(\pi_1(\mathbb{C}P^1 - \{0, 1, \infty\})$ in $PGL(2, \mathbb{C})$ is denoted by Γ, called the monodromy group of the differential equation.

Consider the function $u(z) = \frac{w_1}{w_2}$. Then u is multivalued on $\mathbb{C}P^1 - \{0, 1, \infty\}$, and for each z, the set of the values of $u(z)$ are related under the fractional linear action of $\Gamma = \rho(\pi_1(\mathbb{C}P^1 - \{0, 1, \infty\})$. Denote the image of u in $\mathbb{C}P^1$ by D. Then u induces a well-defined holomorphic map

$$u : \mathbb{C}P^1 - \{0, 1, \infty\} \to \Gamma \backslash D. \tag{6}$$

Schwarz proved that when the parameters $a = b = \frac{1}{12}$, $c = \frac{2}{3}$, for suitable choices of the basepoint z_0 and solutions w_1, w_2 near z_0, the image D is equal to the unit disc $\{z \in \mathbb{C} \mid |z| < 1\} \subset \mathbb{C}P^1$, and Γ is isomorphic to $SL(2, \mathbb{Z})$ (in fact, equal to the image of $SL(2, \mathbb{Z})$ under the Cayley transform $SL(2, \mathbb{C}) \to SU(1, 1)$, where $SU(1, 1)/U(1) = \{z \in \mathbb{C} \mid |z| < 1\}$, and the Cayley transform maps $\mathbf{H}$ biholomorphically onto the unit disc $\{z \in \mathbb{C} \mid |z| < 1\}$). The map u in Equation (6) is injective, and hence $\Gamma \backslash D$ uniformizes the partial compactification $\mathbb{C}P^1 - \{\infty\}$ of $\mathbb{C}P^1 - \{0, 1, \infty\}$.

Such problems have been studied by Gauss and others. One of the motivations is to find new transcendental functions that arise as automorphic functions (or forms) and are important in number theory (see [Ho1] [Ho2] [Yo1] [Yo2]). Another application is to generate interesting discrete subgroups via the monodromy groups (see [DM]). The Picard modular groups discussed below are the monodromy group of certain partial differential equations with regular singularities in 2 complex variables with the unit disc in $\mathbb{C}$ replaced by the unit ball in $\mathbb{C}^2$ (see [Ho1]).

Realization of discrete series.

Let G be a connected semisimple Lie group and $X = G/K$ the associated symmetric space of noncompact type. Since G acts on X, it also acts on spaces of solutions of invariant differential operators on X. In fact, the action of G on X and discrete subgroups $\Gamma \subset G$ acting cocompactly on X can be used to realize and to understand the discrete series representations of G. Recall that an irreducible representation of G is called a discrete series representation if it appears as an irreducible subrepresentation of the regular representation of G

in $L^2(G)$. These discrete series representations can be realized in the space of L^2-solutions of certain elliptic differential equations on X (see [AS] for details).

In the case of $G = SL(2,\mathbb{R})$, the discrete series representations $\mathcal{D}_n^{\pm}$, $n \geq 2$, can be realized as follows. For each integer $n \geq 2$, define

$$V_n = \{f(z) \text{ holomorphic on } \mathbf{H} \mid ||f||^2 = \int_{\mathbf{H}} |f(z)|^2 y^{n-2} dxdy < +\infty\}.$$

Then $SL(2,\mathbb{R})$ acts unitarily on V_n through $\mathcal{D}_n^+$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} f(z) = (-bz+d)^{-n} f(\frac{az-c}{-bz+d}).$$

The representations $\mathcal{D}_n^-$ can also be realized similarly (see [Kn, p. 35] for details).

Modular forms.

Modular forms on $\mathbf{H}$ with respect to $\Gamma = SL(2,\mathbb{Z})$ are holomorphic functions satisfying some transform rule under the action of elements of Γ and bounds at infinity.

Specifically, for any integer $k \geq 0$, a holomorphic function on $\mathbf{H}$ is called a *modular form of weight* k if

$$f(\gamma z) = (cz+d)^{2k} f(z)$$

for all

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}),$$

and the Fourier coefficients of f at infinity,

$$f(z) = \sum a_n e^{2\pi i n z},$$

satisfy $a_n = 0$ for $n < 0$. Each such modular form defines a L-series

$$L(s) = \sum_{n=0}^{\infty} a_n n^{-s}, \quad \mathrm{Re}(s) \gg 0,$$

which enjoy nice properties such as meromorphic continuation to $s \in \mathbb{C}$ and the functional equation. They are important in number theory, string theory and finite group theory (see [CR] and its references). Briefly, the properties of $L(s)$ reflect regularities of the sequence of numbers a_n.

More examples.

The example

$$\Gamma\backslash X = SL(2,\mathbb{Z})\backslash\mathbf{H} = SL(2,\mathbb{Z})\backslash SL(2,\mathbb{R})/SO(2)$$

is particularly important and can be generalized in several directions. The first generalization is to consider

$$G = SL(n, \mathbb{R}), \quad K = SO(n), \quad \Gamma = SL(n, \mathbb{Z}).$$

Then $X = SL(n, \mathbb{R})/SO(n)$ can be identified with the space of positive definite matrices of determinant one. In this case, fundamental domains of $SL(n, \mathbb{Z})$ are more difficult to describe and need infinitely many inequalities. The family $SL(n, \mathbb{R})/SO(n)$ is an important example of so-called linear symmetric spaces.

This method can be generalized by taking Γ to be the subgroup of a linear algebraic (matrix) group consisting of integral matrices and gives rise to arithmetic groups, which will be discussed in §4 below.

Another generalization is to view $\mathbf{H}$ as the Poincare disc and consider bounded symmetric domains. In fact, let $D = \{z \in \mathbb{C} \mid |z| < 1\}$ be the unit disc. Then the map

$$\mathbf{H} \to D, \quad z \mapsto w = \frac{z - i}{z + i},$$

is a biholomorphic map.

The unit disc D is the simplest, and most important of the class of bounded symmetric domains.

Definition 2.19 *A bounded domain Ω in $\mathbb{C}^n$ is called symmetric if for every $z \in \Omega$, there exists an involutive biholomorphic automorphism s_z of Ω such that z is an isolated fixed point of s_z.*

To show that D is a symmetric domain, we note that at the origin $z = 0$, the symmetry $s_0(z) = -z$. Since the group

$$SU(1,1) = \{\begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix} \mid a, b \in \mathbb{C}, |a|^2 - |b|^2 = 1\}$$

acts transitively on D, the conjugates of s_0 give the symmetries at other points s_z.

It is known that any bounded symmetric domain endowed with the Bergman metric is a (Hermitian) symmetric space of noncompact type, and quotients of bounded symmetric domains give rise to Shimura varieties. The Bergman metric of the unit disc D is given by a multiple of

$$ds^2 = \frac{|dz|^2}{(1 - |z|^2)^2}.$$

These two generalizations show that the upper half plane

$$\mathbf{H} = SL(2, \mathbb{R})/SO(2) \cong SU(1,1)/U(1)$$

is an important symmetric space by being the simplest linear symmetric spaces and Hermitian symmetric spaces.

3 Algebraic groups

In this section, we recall basic facts about linear algebraic groups in order to define arithmetic subgroups and discuss reduction theories in the latter sections. The basic references are [Bo2] [Bo4].

Definition 3.1 *A variety* $\mathbf{G}$ *over* $\mathbb{C}$ *is called an algebraic group if it is also a group and the group operations*

$$\mu : \mathbf{G} \times \mathbf{G} \to \mathbf{G}, \quad (g_1, g_2) \mapsto g_1 g_2,$$

$$i : \mathbf{G} \to \mathbf{G}, \quad g \mapsto g^{-1}$$

are morphisms of varieties.

There are two particularly important types of algebraic groups depending on whether $\mathbf{G}$ is a complete variety or an affine variety. If $\mathbf{G}$ is a complete variety, then $\mathbf{G}$ is an abelian variety (see [GH, p. 325]). We will be mainly interested in affine algebraic groups, which are equivalent to linear algebraic (or matrix) groups. Specifically, a linear algebraic group $\mathbf{G}$ is a Zariski closed subgroup of some general linear group $GL(n, \mathbb{C})$:

$$\mathbf{G} = \{g = (g_{ij}) \in GL(n, \mathbb{C}) \mid P_a(g_{ij}) = 0, a \in A\},$$

where each P_a is a polynomial in g_{ij}, and A a parameter space.

The first example of linear algebraic group is $GL(n, \mathbb{C})$. It is contained in the affine space of $n \times n$-matrices $M_{n\times n}(\mathbb{C}) \cong \mathbb{C}^{n^2}$ but this embedding does not realize $GL(n, \mathbb{C})$ as an affine variety. Instead we use the embedding

$$GL(n, \mathbb{C}) \to M_{n\times n}(\mathbb{C}) \times \mathbb{C} = \mathbb{C}^{n^2+1}, \quad (g_{ij}) \mapsto ((g_{ij}), (\det(g_{ij}))^{-1}).$$

Let X_{ij}, Z be the coordinates of $M_{n\times n}(\mathbb{C}) \times \mathbb{C}$. Then the image is the affine hypersurface defined by

$$\det(X_{ij}) Z = 1,$$

which is clearly a polynomial in X_{ij}, Z with coefficients in $\mathbb{Z}$. It can be checked easily that the group operations on $GL(n, \mathbb{C})$ are given by polynomials in X_{ij} and Z. Hence $GL(n, \mathbb{C})$ is an affine algebraic variety as defined above.

Linear algebraic groups often occur as the automorphism group of some structures such as determinant, quadratic forms. For example,

$$SL(n, \mathbb{C}) = \{g \in GL(n, \mathbb{C}) \mid \det g = 1\};$$

$$Sp(n, \mathbb{C}) = \{g \in GL(2n, \mathbb{C}) \mid \det g = 1, \omega(gX, gY) = \omega(X, Y), X, Y \in \mathbb{C}^{2n}\},$$

where

$$\omega(X, Y) = x_1 y_{2n} + x_2 y_{2n-1} + \cdots + x_n y_{n+1} - x_{n+1} y_n - \cdots - x_{2n} y_1$$

is a skew-symmetric form; and

$$SO(2n,\mathbb{C}) = \{g \in GL(2n,\mathbb{C}) \mid \det g = 1, \langle gX, gY\rangle = \langle X, Y\rangle, X, Y \in \mathbb{C}^{2n}\},$$

where

$$\langle X, Y\rangle = x_1 y_{2n} + \cdots + x_{2n} y_1$$

a symmetric quadratic form. More generally, let F be a non-degenerate quadratic form on $\mathbb{C}^n$ with the corresponding symmetric matrix A. Define

$$O(F) = \{g \in GL(n,\mathbb{C}) \mid {}^t g A g = A\} = \{g \in GL(n,\mathbb{C}) \mid F(g{\cdot}v) = F(v),\ v \in \mathbb{C}^n\},$$

called the orthogonal group of F. Clearly they are affine algebraic varieties. As seen from these examples, the polynomial equations defining the algebraic groups arise from additional structures on the vector space $\mathbb{C}^n$ that need to be preserved by the groups. For example, $SL(n,\mathbb{C})$ can also be defined as the subgroup of $GL(n,\mathbb{C})$ which preserves the form $dX_{11} \wedge dX_{12} \wedge \cdots \wedge dX_{nn}$.

In the above discussions, we have not specified a field of definition. A linear algebraic group $\mathbf{G} \subset GL(n,\mathbb{C})$ is said to be defined over $\mathbb{Q}$ if the polynomials which define $\mathbf{G}$ as a subvariety have coefficients in $\mathbb{Q}$. It is clear that all the examples above are defined over $\mathbb{Q}$. In these notes, we will assume that algebraic groups are defined over $\mathbb{Q}$ unless indicated otherwise.

A linear algebraic group $\mathbf{T}$ is called an algebraic torus if it is isomorphic to a product of $\mathbb{C}^* = GL(1,\mathbb{C})$. If the isomorphism is defined over $\mathbb{Q}$ (resp. $\mathbb{R}$), the torus $\mathbf{T}$ is said to split over $\mathbb{Q}$ (resp. $\mathbb{R}$).

Consider the algebraic group

$$\mathbf{T}_1 = \{g \in SL(2,\mathbb{C}) \mid {}^t g \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} g = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\}.$$

It can be checked easily that if $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{T}_1$, then $b = d = 0$, $c = a^{-1}$, and hence $g = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$. This implies that $\mathbf{T}_1$ is isomorphic to $GL(1,\mathbb{C})$ over $\mathbb{Q}$ under the map $g \to a$, and hence $\mathbf{T}_1$ splits over $\mathbb{Q}$.

On the other hand, the algebraic group

$$\mathbf{T}_2 = \{g \in SL(2,\mathbb{C}) \mid {}^t g \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} g = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\}$$

is also a torus defined over $\mathbb{Q}$ but does not split over $\mathbb{Q}$ or $\mathbb{R}$. In fact, the real locus $\mathbf{T}_2(\mathbb{R}) = SO(2,\mathbb{R})$ is compact. To see that $\mathbf{T}_2$ is a torus, we note that $\mathbf{T}_2$ preserves the quadratic form

$$Q(X, X) = x_1^2 + x_2^2,$$

while $\mathbf{T}_1$ preserves the quadratic form

$$\langle X, X\rangle = 2x_1 x_2,$$

and these two forms are equivalent over $\mathbb{C}$, i.e., the quadratic form $x_1^2+x_2^2$ splits as $(x_1+ix_2)(x_1-ix_2)$ over $\mathbb{C}$.

Consider another algebraic group defined over $\mathbb{Q}$,

$$\mathbf{T}_3 = \{g \in SL(2,\mathbb{C}) \mid {}^t g \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} g = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}\}.$$

Since the quadratic form preserved by $\mathbf{T}_3$ is $x_1^2-2x_2^2$ which splits over $\mathbb{R}$ as $(x_1+\sqrt{2}x_2)(x_1-\sqrt{2}x_2)$ but not over $\mathbb{Q}$, it can be shown that $\mathbf{T}_3$ splits over $\mathbb{R}$ but not over $\mathbb{Q}$.

A linear algebraic group $\mathbf{G}$ is called unipotent if every element g of $\mathbf{G}$ is unipotent, i.e., $(g-I)^k=0$ for some integer k. For example, the additive group

$$\mathbf{G}_a = \{\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \mid b \in \mathbb{C}\}$$

is unipotent. Note that $\mathbf{G}_a \cong \mathbb{C}$, but we need this realization to embed it into $GL(n,\mathbb{C})$ as a linear algebraic group.

More generally, the subgroup $\mathbf{U}$ of $GL(n,\mathbb{C})$ consisting of upper triangular matrices with ones on the diagonal is unipotent. Clearly, any subgroup of $\mathbf{U}$ is unipotent as well. The converse is also true, i.e., any connected unipotent algebraic group is isomorphic to a subgroup of $\mathbf{U}$.

A linear algebraic group is called solvable if it is solvable as an abstract group, i.e., the derived series terminates, $G = G^{(0)} \supset G^{(1)} \supset \cdots \supset G^{(l)} = \{e\}$ for some l, where $G^{(i)} = [G^{(i)}, G^{(i)}]$.

It can be checked easily that the subgroup $\mathbf{B}$ of $GL(n,\mathbb{C})$ of upper triangular matrices is solvable. Hence, the above discussions show that a unipotent group is always solvable. On the other hand, for a solvable algebraic group $\mathbf{G}$ defined over $\mathbb{Q}$, let $\mathbf{U}$ be its normal subgroup consisting of all the unipotent elements. Then there exists a maximal torus $\mathbf{T}$ defined over $\mathbb{Q}$ such that $\mathbf{G}$ is the semi-direct product of $\mathbf{T}$ and $\mathbf{U}$.

The radical $\mathbf{R}(\mathbf{G})$ of an algebraic group $\mathbf{G}$ is the maximal connected normal solvable subgroup of $\mathbf{G}$, and the unipotent radical $\mathbf{R}_{\mathbf{U}}(\mathbf{G})$ is the maximal connected unipotent normal subgroup of $\mathbf{G}$. If $\mathbf{G}$ is defined over $\mathbb{Q}$, then $\mathbf{R}(\mathbf{G}), \mathbf{R}_{\mathbf{U}}(\mathbf{G})$ are also defined over $\mathbb{Q}$. A linear algebraic group $\mathbf{G}$ is called semisimple if the radical $\mathbf{R}(\mathbf{G}) = \{e\}$, and reductive if the unipotent radical $\mathbf{R}_{\mathbf{U}}(\mathbf{G}) = \{e\}$.

Clearly, $\mathbf{G}/\mathbf{R}(\mathbf{G})$ is semisimple and $\mathbf{G}/\mathbf{R}_{\mathbf{U}}(\mathbf{G})$ is reductive. It is known that if $\mathbf{G}$ is defined over $\mathbb{Q}$, then there exists a maximal reductive group $\mathbf{H}$ defined over $\mathbb{Q}$ such that $\mathbf{G} = \mathbf{H} \cdot \mathbf{R}_{\mathbf{U}}(\mathbf{G})$.

Though we are mainly interested in semisimple linear algebraic groups, reductive groups occur naturally when we consider parabolic subgroups and boundary components of compactifications of locally symmetric spaces. If $\mathbf{G}$ is a connected reductive algebraic group, then the derived subgroup $\mathbf{G}' = [\mathbf{G}, \mathbf{G}]$ is semisimple, and there exists a central torus $\mathbf{T}$ such that $\mathbf{G} = \mathbf{T} \cdot \mathbf{G}'$.

For an algebraic group defined over $\mathbb{Q}$, an important notion is its $\mathbb{Q}$-rank, which plays a fundamental role in the geometry at infinity of locally symmetric spaces.

Let $\mathbf{G}$ be a connected reductive linear algebraic group. Then all the maximal tori of $\mathbf{G}$ are conjugate, and the common dimension is called the absolute (or $\mathbb{C}$) rank of $\mathbf{G}$, denoted by $rk_{\mathbb{C}}(\mathbf{G})$. If $\mathbf{G}$ is defined over $\mathbb{Q}$, then all the maximal $\mathbb{Q}$-split tori of $\mathbf{G}$ are conjugate over $\mathbb{Q}$, i.e., by elements of $\mathbf{G}(\mathbb{Q})$, and the common dimension is called the $\mathbb{Q}$-rank of $\mathbf{G}$, denoted by $rk_{\mathbb{Q}}(\mathbf{G})$. Similarly, the common dimension of maximal $\mathbb{R}$-split tori is called the $\mathbb{R}$-rank of $\mathbf{G}$, denoted by $rk_{\mathbb{R}}(\mathbf{G})$. The examples of tori show that these ranks are in general not equal to each other. If $\mathbb{C}$-rank is equal to the $\mathbb{Q}$-rank, $\mathbf{G}$ is said to split over $\mathbb{Q}$. For example, when $\mathbf{G} = SL(n, \mathbb{C})$, $rk_Q(\mathbf{G}) = rk_{\mathbb{R}}(\mathbf{G}) = rk_{\mathbb{C}}(\mathbf{G}) = n - 1$, and $SL(n, \mathbb{C})$ splits over $\mathbb{Q}$.

Proposition 3.2 *If F is a non-degenerate quadratic form on a $\mathbb{Q}$-vector space V with coefficients in $\mathbb{Q}$, then the orthogonal group $O(F)$ of F has positive $\mathbb{Q}$-rank if and only if F represents 0 over $\mathbb{Q}$, i.e., $F = 0$ has a nontrivial solution over $\mathbb{Q}$.*

Proof. If the $\mathbb{Q}$-rank of $O(F)$ is positive, then there exists a nontrivial split torus $\mathbf{T}$ over $\mathbb{Q}$ in $O(F)$. Since $\mathbf{T}$ splits over $\mathbb{Q}$, we can diagonalize it, or equivalently we can decompose $\mathbb{C}^n$ as a direct sum of weight spaces

$$\mathbb{C}^n = \oplus_\mu V_\mu,$$

where $\mathbf{T}$ acts on V_μ according to the character μ: for $v \in V_\mu$, $t \in \mathbf{T}$, $t \cdot v = t^\mu v$. Since $\mathbf{T}$ acts nontrivially on $\mathbb{C}^n$, there exists a nontrivial weight μ_0 in the above decomposition. Take a nonzero vector $v \in V_{\mu_0}$. Then by definition of $O(F)$, for any $t \in \mathbf{T}$, $F(t \cdot v) = F(v)$. On the other hand,

$$F(t \cdot v) = F(t^{\mu_0} v) = t^{2\mu_0} F(v).$$

Since $2\mu_0$ is nontrivial, there exists $t \in \mathbf{T}$ with $t^{2\mu_0} \neq 1$ and hence $F(v) = 0$.

Conversely, suppose there exists $v \in \mathbb{C}^n$, $v \neq 0$, such that $F(v) = 0$. Since F is non-degenerate, there exists $v_2 \in \mathbb{C}^n$ such that $F(v, v_2) = 1$. The orthogonal complement of the subspace spanned by $v_1 = v$ and v_2 with respect to F has dimension $n - 2$. Let $v_3, \cdots, v_n$ be a basis of this subspace, and $x_1, \cdots, x_n$ be the coordinates of $\mathbb{C}^n$ with respect to the basis $v_1, v_2, \cdots, v_n$. Then the quadratic form F can be written as

$$F(x_1, \cdots, x_n) = x_1 x_2 + F'(x_3, \cdots, x_n),$$

where F' is a quadratic form on the complement subspace. Clearly, the action of $\mathbf{T} = \mathbb{C}^\times$ on $\mathbb{C}^n$ by

$$x_1 \mapsto t x_1,\ x_2 \mapsto t^{-1} x_2,\ x_3 \mapsto x_3,\ \cdots, x_n \mapsto x_n,$$

preserves the form F and hence the $\mathbb{Q}$-split torus $\mathbf{T}$ is contained in $O(F)$, i.e., the $\mathbb{Q}$-rank of $O(F)$ is positive.

Let $\mathbf{G}$ be a connected linear algebraic group. Then a closed subgroup $\mathbf{P}$ of $\mathbf{G}$ is called a *parabolic subgroup* if $\mathbf{G}/\mathbf{P}$ a projective variety, which is equivalent to that $\mathbf{P}$ contains a maximal connected solvable subgroup of $\mathbf{G}$, i.e., a *Borel subgroup* of $\mathbf{G}$. These conditions are also equivalent to that $\mathbf{G}/\mathbf{P}$ is compact. Parabolic subgroups are important for the purpose to understand the geometry at infinity of symmetric and locally symmetric spaces.

Assume for the rest of this paper that $\mathbf{G}$ is a linear algebraic group defined over $\mathbb{Q}$. If a parabolic subgroup $\mathbf{P}$ is defined over $\mathbb{Q}$, it is called a rational parabolic subgroup. It is known that minimal rational parabolic subgroups of $\mathbf{G}$ are conjugate over $\mathbb{Q}$. For each fixed minimal rational parabolic subgroup $\mathbf{P}$, there are only finitely many rational parabolic subgroups containing it, called the standard parabolic subgroups and can be described explicitly in terms of $\mathbf{P}$ and the associated roots.

We illustrate the above definitions and concepts through the example of $\mathbf{G} = GL(n, \mathbb{C})$. The group $\mathbf{G}$ is reductive but not semisimple. In fact, the center $\mathcal{Z}(\mathbf{G}) = \mathbb{C}^\times Id \cong GL(1)$ and hence $\mathbf{G}$ has the nontrivial radical $\mathbf{R}(\mathbf{G})$. The derived group $\mathbf{G}' = SL(n, \mathbb{C})$ is semisimple and $\mathbf{G} = \mathcal{Z}(\mathbf{G})\mathbf{G}'$. The torus $\mathbf{T}$ consisting of diagonal matrices

$$\mathbf{T} = \{diag(t_1, \cdots, t_n) \mid t_1, \cdots, t_n \in \mathbb{C}^\times\} \cong (\mathbb{C}^\times)^n$$

is a maximal $\mathbb{Q}$-split torus of $\mathbf{G}$ (in fact, it is also a maximal torus over $\mathbb{C}$ in $\mathbf{G}$), and hence the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to n.

The subgroup $\mathbf{B}$ of upper triangular matrices is a minimal rational parabolic subgroup of $\mathbf{G}$ and is also a Borel subgroup, and the rational parabolic subgroups containing $\mathbf{B}$ are given by groups of block upper triangular matrices. In this case, these standard parabolic subgroups are stabilizers of flags, i.e., a sequence of increasing subspaces: $V_0 = 0 \subset V_1 \subset \cdots V_k = \mathbb{C}^n$. The minimal parabolic subgroup $\mathbf{B}$ is the stabilizer of the full flag with $V_i = \mathbb{C}^i$, and the maximal parabolic subgroups are the stabilizers of the nontrivial flags with minimal length $k = 2$.

4 Arithmetic subgroups

In this section we introduce arithmetic groups, study their basic properties and conclude with some important examples of arithmetic groups. The basic reference of this section is [Bo2].

Let $\mathbf{G} \subset GL(n, \mathbb{C})$ be a linear algebraic group defined over $\mathbb{Q}$, not necessarily reductive. Let $\mathbf{G}(\mathbb{Q}) \subset GL(n, \mathbb{Q})$ be the set of its rational points, and $\mathbf{G}(\mathbb{Z}) \subset GL(n, \mathbb{Z})$ the set of its elements with integral entries, which can be identified with the stabilizer of the standard lattice $\mathbb{Z}^n$ in $\mathbb{R}^n$.

Definition 4.1 *A subgroup $\Gamma \subset \mathbf{G}(\mathbb{Q})$ is called an arithmetic subgroup if it is commensurable to $\mathbf{G}(\mathbb{Z})$, i.e., $\Gamma \cap \mathbf{G}(\mathbb{Z})$ has finite index in both Γ and $\mathbf{G}(\mathbb{Z})$.*

As an abstract affine algebraic group defined over $\mathbb{Q}$, $\mathbf{G}$ admits different embeddings into $GL(n', \mathbb{C})$, where n' might be different from n. The above definition depends on the embedding $\mathbf{G} \subset GL(n, \mathbb{C})$ and the integral subgroup $GL(n, \mathbb{Z})$. If we choose a different embedding, for example using a basis of $\mathbb{C}^n$ over $\mathbb{Q}$ different from the standard basis $e_1 = (1, 0, \cdots, 0), \cdots, e_n = (0, \cdots, 0, 1)$, then we will get a different integral subgroup $\mathbf{G}(\mathbb{Z})$ of $GL(n, \mathbb{C})$ defined with respect to this basis.

It turns out that these different embeddings $\mathbf{G} \subset GL(n', \mathbb{C})$ and different choices of integral structures lead to the same class of arithmetic groups.

Proposition 4.2 *Let $\mathbf{G}, \mathbf{G}'$ be two linear algebraic groups defined over $\mathbb{Q}$, and $\varphi : \mathbf{G} \to \mathbf{G}'$ an isomorphism defined over $\mathbb{Q}$. Then $\varphi(\mathbf{G}(\mathbb{Z}))$ is commensurable to $\mathbf{G}'(\mathbb{Z})$.*

Proof. Since

$$[\mathbf{G}'(\mathbb{Z}) : \mathbf{G}'(\mathbb{Z}) \cap \varphi(\mathbf{G}(\mathbb{Z}))] = [\varphi^{-1}(\mathbf{G}'(\mathbb{Z})) : \varphi^{-1}(\mathbf{G}'(\mathbb{Z})) \cap \mathbf{G}(\mathbb{Z})],$$

it suffices to show that

$$[\varphi(\mathbf{G}(\mathbb{Z})) : \varphi(\mathbf{G}(\mathbb{Z})) \cap \mathbf{G}'(\mathbb{Z})] < +\infty, \quad [\varphi^{-1}(\mathbf{G}'(\mathbb{Z})) : \varphi^{-1}(\mathbf{G}'(\mathbb{Z})) \cap \mathbf{G}(\mathbb{Z})] < +\infty.$$

Since the arguments are the same for both inequalities, we will only prove the first one. The idea is to find a subgroup $\Gamma' \subset \mathbf{G}(\mathbb{Z})$ of finite index such that $\varphi(\Gamma') \subset \mathbf{G}'(\mathbb{Z})$.

By realizing $\mathbf{G}, \mathbf{G}'$ as affine varieties in $M_{m\times m}(\mathbb{C})$ and $M_{n\times n}(\mathbb{C})$ respectively, we can assume that the morphism φ is given by

$$\varphi((x_{ij})) = (\varphi_{kl}(x_{11}, \cdots, x_{nn})),$$

where $\varphi_{kl}(x_{11}, \cdots, x_{nn})$ is a polynomial in $x_{11}, \cdots, x_{nn}$ with rational coefficients. Introduce new variables $y_{ij} = x_{ij} - \delta_{ij}$, and put

$$\psi_{kl}(y_{11}, \cdots, y_{nn}) = \varphi_{kl}(x_{11}, \cdots, x_{nn}) - \delta_{kl}.$$

Since $\varphi(Id) = Id$, $\psi_{kl}(0, \cdots, 0) = 0$ and hence ψ_{kl} has zero constant term.

Since the coefficients φ_{kl} are rational, we can pick an integer d such that for all k, l, $d\psi_{kl}$ has integral coefficients. Let Γ' be the congruence subgroup of $\mathbf{G}(\mathbb{Z})$ of level d,

$$\Gamma' = \{g \in \mathbf{G}(\mathbb{Z}) \mid g \equiv Id \mod d\}.$$

Clearly, $[\mathbf{G}(\mathbb{Z}) : \Gamma'] \leq |GL(n, \mathbb{Z}/d\mathbb{Z})| < +\infty$ and hence Γ' is of finite index. Then for $g \in \Gamma'$, $\psi(g - Id)$ is integral since $d | g - Id$. This implies that

$$\varphi(g) = \psi(g - Id) + Id \in \mathbf{G}'(\mathbb{Z}).$$

Therefore $\varphi(\Gamma') \subset \mathbf{G}'(\mathbb{Z})$.

Corollary 4.3 *If Γ is an arithmetic subgroup of $\mathbf{G}$, then for any $g \in \mathbf{G}(\mathbb{Q})$, $g\Gamma g^{-1}$ is also an arithmetic subgroup.*

To discuss the Hilbert modular groups, the Bianchi groups and Picard modular groups below, we need a slightly more general set-up for arithmetic groups. Let F be a number field, i.e., a finite extension of $\mathbb{Q}$, and $\mathcal{O}_F$ its ring of integers. Let $\mathbf{G} \subset GL(n, \mathbb{C})$ be a linear algebraic group defined over F. A subgroup Γ of $\mathbf{G}(F)$ is called arithmetic if it is commensurable to $G(\mathcal{O}_F) = \mathbf{G} \cap GL(n, \mathcal{O}_F)$. It turns out that such an arithmetic subgroup is also an arithmetic subgroup according to the previous definition and hence we do not get more arithmetic subgroups by considering general number fields. On the other hand, it is often convenient to use some naturally occuring number fields to define arithmetic groups. In fact, by the functor of restriction of scalars, there is an algebraic group $Res_{F/\mathbb{Q}}\mathbf{G}$ defined over $\mathbb{Q}$ such that $Res_{K/\mathbb{Q}}\mathbf{G}(\mathbb{Q}) = \mathbf{G}(F)$, and $\mathbf{G}(\mathcal{O}_F)$ is commensurable to $Res_{K/\mathbb{Q}}\mathbf{G}(\mathbb{Z})$ under this identification.

The functor of restriction scalars is defined as follows (see [PR] for more details). For any $x \in F$, the multiplication

$$x : F \to F, \quad y \mapsto xy,$$

is a $\mathbb{Q}$-linear transformation of the vector space F over $\mathbb{Q}$. Let $v_1, \cdots, v_d$ be a basis of F over $\mathbb{Q}$ considered as a vector space over $\mathbb{Q}$. Then each element $x \in F$ corresponds to a matrix in $M_{d\times d}(\mathbb{Q})$, and hence we get a faithful representation

$$\rho : F \to M_{d\times d}(\mathbb{Q}).$$

Clearly $\rho(F)$ is a linear subspace of $\mathbb{Q}^d$. Let $f_1 = 0, \cdots, f_r = 0$ be the linear equations with coefficients in $\mathbb{Q}$ defining $\rho(F)$:

$$\rho(F) = \{(y^{\alpha\beta}) \in M_{d\times d}(\mathbb{Q}) \mid f_1(y^{\alpha\beta}) = \cdots = f_r(y^{\alpha\beta}) = 0\}.$$

Suppose the linear algebraic group $\mathbf{G}$ is defined by

$$\mathbf{G} = \{g \in GL(n, \mathbb{C}) \mid P_\ell(g) = 0, \ell \in I\},$$

where P_ℓ are polynomials in g_{ij}, $g = (g_{ij})$, with coefficients in F. For each polynomial

$$P_\ell(g) = \sum_{\gamma_{11}, \cdots, \gamma_{nn}} a_{\gamma_{11}\cdots\gamma_{nn}} g_{11}^{\gamma_{11}} \cdots g_{nn}^{\gamma_{nn}}, \quad a_{\gamma_{11}\cdots\gamma_{nn}} \in F,$$

define a polynomial $\tilde{P}(y)$ with coefficients in $M_{d\times d}(\mathbb{Q})$,

$$\tilde{P}_\ell(y) = \sum_{\gamma_{11}, \cdots, \gamma_{nn}} \rho(a_{\gamma_{11}\cdots\gamma_{nn}})(y_{11}^{\alpha\beta})^{\gamma_{11}} \cdots (y_{nn}^{\alpha\beta})^{\gamma_{nn}},$$

where for each entry g_{ij} of g, there is a matrix $(y_{ij}^{\alpha\beta}) \in M_{d\times d}$ of variables. Then $Res_{F/\mathbb{Q}}\mathbf{G}$ is a linear subgroup of $GL(nd, \mathbb{C}) \subset M_{n\times n}(M_{d\times d}(\mathbb{C})) = M_{nd\times nd}(\mathbb{C})$ defined by

$$\begin{aligned} Res_{F/\mathbb{Q}}\mathbf{G} = &\{(y_{11}^{\alpha\beta}), \cdots, (y_{nn}^{\alpha\beta}) \mid (1) \text{ for every pair } \ i, j, \quad f_1(y_{ij}^{\alpha\beta}) = \cdots = f_r(y_{ij}^{\alpha\beta}) = 0, \\ &(2)\ \tilde{P}_\ell(y) = 0, \ell \in I\}, \end{aligned}$$

where all the polynomials clearly have coefficients in $\mathbb{Q}$.

One can see that $Res_{F/\mathbb{Q}}\mathbf{G}(\mathbb{Q}) = \mathbf{G}(F)$. In fact, the conditions in (1) guarantee that $(y_{11}^{\alpha\beta}), \cdots, (y_{nn}^{\alpha\beta})$ belong to $\rho(F) \subset M_{d\times d}(\mathbb{Q})$, and the conditions in (2) are carried from the defining polynomials of $\mathbf{G}$. The $\mathbb{Q}$-rank of $Res_{F/\mathbb{Q}}\mathbf{G}$ is equal to the F-rank of $\mathbf{G}$.

For a linear algebraic group $\mathbf{G}$ defined over $\mathbb{Q}$, any arithmetic group $\Gamma \subset \mathbf{G}(\mathbb{Q})$ is a discrete subgroup of $\mathbf{G}(\mathbb{R})$, which is obtained by the canonical embedding $\mathbb{Q} \hookrightarrow \mathbb{R}$. On the other hand, for an algebraic group $\mathbf{G}$ defined over a number field F and an arithmetic subgroup $\Gamma \subset \mathbf{G}(F)$, the natural Lie group containing Γ as a discrete subgroup is

$$\prod_{i=1}^{s+t} \mathbf{G}(F_{\mu_i}),$$

where $\mu_1, \cdots, \mu_s$ are all the different real embeddings $F \hookrightarrow \mathbb{R}$, and $\mu_{s+1}, \overline{\mu_{s+1}}, \cdots, \mu_{s+t}, \overline{\mu_{s+t}}$ are the different complex embeddings $K \hookrightarrow \mathbb{C}$. In fact,

$$Res_{F/\mathbb{Q}}\mathbf{G}(\mathbb{R}) \cong \prod_{i=1}^{s+t} \mathbf{G}(F_{\mu_i}).$$

If $s+t > 1$, the image of $\mathbf{G}(\mathcal{O}_F)$ in each factor $\mathbf{G}(F_{\mu_i})$ is not discrete in general. For example, let d be a positive square free integer and $F = \mathbb{Q}(\sqrt{d})$. Then $s = 2$ and $t = 0$. Under each real embedding, the image of $GL(2, \mathcal{O}_F)$ in $GL(2, \mathbb{R})$ is not discrete.

We illustrate the above constructions by a simple example. Let $\mathbf{G}_a$ be the additive group group, which is defined over $\mathbb{Q}$ and hence also defined over any number field F. Then $\mathbf{G}_a(\mathcal{O}_F) = \mathcal{O}_F$. The algebraic group $Res_{F/\mathbb{Q}}\mathbf{G}_a$ is isomorphic to the commutant of $\rho(F)$ in $M_{d\times d}(\mathbb{C})$, and the image of $\mathbf{G}_a(\mathcal{O}_F)$ in

$$Res_{F/\mathbb{Q}}\mathbf{G}(\mathbb{R}) = \prod_{i=1}^{s+t} \mathbf{G}(F_{\mu_i}) = (\mathbb{R})^s \times (\mathbb{C})^t = \mathbb{R}^{s+2t}$$

is a lattice. Another natural algebraic group associated with F is the multiplicative group of elements of $F^\times$ of norm 1, i.e., the subgroup of $Res_{F/\mathbb{Q}}GL(1)$ defined by $Nr(x) = 1$, where Nr is the norm of F over $\mathbb{Q}$. The group of units of $\mathcal{O}_F$ is a discrete subgroup of this algebraic group.

In the above discussions, we chose an integral structure on a vector space over $\mathbb{Q}$ by fixing a basis over $\mathbb{Q}$. Another important point of view is to choose a lattice compatible with the rational structure. Let V be a vector space over $\mathbb{Q}$, and L a lattice in V. Define

$$GL(V, L) = \{g \in GL(V) \mid gL = L\}.$$

For any algebraic group $\mathbf{G} \subset GL(V \otimes_{\mathbb{Q}} \mathbb{C})$ defined over $\mathbb{Q}$, define the subgroup $\mathbf{G}(L)$ of L-units by

$$\mathbf{G}(L) = \mathbf{G}(\mathbb{Q}) \cap GL(V, L).$$

Then a subgroup Γ of $\mathbf{G}(\mathbb{Q})$ is arithmetic if and only if it is commensurable with $\mathbf{G}(L)$. In the earlier definition, we have chosen a basis of V and take L to be the standard lattice $L = \mathbb{Z}^n$ generated by the basis. For a vector space V over a number field F, and an $\mathcal{O}_F$-module L in V of rank equal to $\dim_F V$, we can similarly define $GL(V, L)$ and the subgroup of L-units $\mathbf{G}(L)$. We apply this method to the following case. Let $\mathbf{A}$ be an algebra over $\mathbb{Q}$ of finite dimension. Let $\mathbf{G}$ be the group of invertible elements in $\mathbf{A}(\mathbb{C})$. If we embed $\mathbf{A}$ into the group $\mathrm{End}(\mathbf{A})$ of linear transformations of $\mathbf{A}$ considered as a vector space over $\mathbb{Q}$, then $\mathbf{G} = \mathbf{A} \cap GL(\mathbf{A} \otimes \mathbb{C})$. Let L be a lattice in $\mathbf{A}$. Define the subgroup of L-units

$$\mathbf{G}(L) = \{g \in \mathbf{A} \mid gL = L\}.$$

Then any subgroup of $\mathbf{A}$ commensurable with $\mathbf{G}(L)$ is an arithmetic subgroup of $\mathbf{G}$.

In the rest of this section, we consider several important arithmetic subgroups associated with lower dimensional symmetric spaces. They are arithmetic Fuchsian groups, the Hilbert modular groups, the Bianchi group, and the Picard modular groups.

Arithmetic Fuchsian groups

We start with the arithmetic Fuchsian groups, which are constructed by using quaternion algebras. Recall that a quaternion algebra A over $\mathbb{Q}$ is a central simple algebra over $\mathbb{Q}$ of dimension 4, i.e., (1) A has no nontrivial two sided ideal, (2) its center is equal to $\mathbb{Q}$.

For any two non-zero elements $a, b \in \mathbb{Q}$, there is a quaternion algebra $\mathcal{H}(a, b)$ defined as the 4-dimensional vector space over $\mathbb{Q}$ with a basis $1, i, j, k$ such that it is an algebra over $\mathbb{Q}$ with multiplication determined by

$$i^2 = a, \quad j^2 = b, \quad ij = -ji = k.$$

When $a = b = -1$, we get the usual quaternion algebra.

The algebra $\mathcal{H}(a, b)$ can be embedded into $M_{2\times 2}(\mathbb{Q}(\sqrt{a}))$ by

$$\rho : x = x_0 + ix_1 + jx_2 + kx_3 \mapsto \begin{pmatrix} x_0 + x_1\sqrt{a} & x_2 + x_3\sqrt{a} \\ b(x_2 - x_3\sqrt{a}) & x_0 - x_1\sqrt{a} \end{pmatrix}. \tag{7}$$

This embedding is obtained by considering $\mathcal{H}(a, b)$ as a two dimensional vector space over $\mathbb{Q}(\sqrt{a}) = \mathbb{Q}(i)$, and $\rho(x)$ is the matrix of the left multiplication by x with respect to the basis $1, j$. Of course, we can also embed $\mathcal{H}(a, b)$ into $M_{4\times 4}(\mathbb{Q})$ by the regular representation. If a is the square of some elements in $\mathbb{Q}^\times$, then $\mathcal{H}(a, b) = M_{2\times 2}(\mathbb{Q})$ and hence is not a division algebra. (Recall that an algebra is called a division algebra if every nonzero element is invertible).

Let ι be the standard involution on $\mathcal{H}(a, b)$:

$$\iota(x_0 + ix_1 + x_2 j + x_3 k) = x_0 - x_1 i - x_2 j - x_3 k.$$

Define

$$Tr(x) = x + \iota(x) = 2x_0, \quad Nr(x) = x\iota(x) = x_0^2 - ax_1^2 - bx_2^2 + abx_3^2.$$

Lemma 4.4 *The algebra $\mathcal{H}(a,b)$ is a division algebra if and only if $Nr(x) \neq 0$ for all $x \neq 0$.*

Proof. Since $Nr(xy) = Nr(x)Nr(y)$, it is clear that if $\mathcal{H}(a,b)$ is a division algebra, then $Nr(x) \neq 0$ for $x \neq 0$. The other direction follows from $Nr(x) = x\iota(x)$.

Proposition 4.5 *The algebra $\mathcal{H}(a,b)$ is either isomorphic to $M_{2\times 2}(\mathbb{Q})$ or is a division algebra.*

Proof. It suffices to prove that if $\mathcal{H}(a,b)$ is not a division algebra, then there exists an isomorphism $\mathcal{H}(a,b) \cong M_{2\times 2}(\mathbb{Q})$. If $\sqrt{a} \in \mathbb{Q}$, it was observed earlier that $\mathcal{H}(a,b) \cong M_{2\times 2}(\mathbb{Q})$. Otherwise, $F = \mathbb{Q}(\sqrt{a}) = \mathbb{Q}(i)$ is a quadratic extension of $\mathbb{Q}$. By assumption, $\mathcal{H}(a,b)$ is not a division algebra, and hence by the above lemma, there exists a nonzero element $x = x_0 + ix_1 + jx_2 + kx_3$ such that $Nr(x) = 0$. Let $n(x_0 + ix_1) = x_0^2 - ax_1^2$ be the norm of F over $\mathbb{Q}$, which is not zero for any nonzero element $x_0 + ix_1$ in F. This implies that $jx_2 + kx_3 \neq 0$, or equivalently $x_2 + ix_3 \neq 0$. Let

$$q_0 + iq_1 = \frac{x_0 + ix_1}{x_2 + ix_3},$$

where $q_0, q_1 \in \mathbb{Q}$. Then

$$Nr(x) = x_0^2 - ax_1^2 - bx_2^2 + abx_3^2 = n(x_0 + ix_1) - bn(x_2 + ix_3),$$

and

$$b = \frac{n(x_0 + ix_1)}{n(x_2 + ix_3)} = n(q_0 + q_1) = q_0^2 - aq_1^2.$$

Define a map $\mathcal{H}(a,b) \to M_{2\times 2}(\mathbb{Q})$ by

$$1 \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i \to \begin{pmatrix} 0 & 1 \\ a & 0 \end{pmatrix}, \quad j \to \begin{pmatrix} q_0 & -q_1 \\ q_1 a & -q_0 \end{pmatrix}, \quad k \to \begin{pmatrix} q_1 a & -q_0 \\ aq_0 & -aq_1 \end{pmatrix}.$$

It can be checked easily that this is an isomorphism.

We can get examples of division algebras using the next result.

Proposition 4.6 *Let b be a prime number, a be any quadratic non-residue* mod b. *Then $\mathcal{H}(a,b)$ is a division algebra.*

Proof. If not, there exists a nonzero element $x = x_0 + ix_1 + jx_2 + kx_3$ such that $Nr(x) = x_0^2 - ax_1^2 - bx_2^2 + abx_3^2 = 0$. We can assume that x_0, x_1, x_2, x_3 are integers and have no common (nontrivial) divisors. Then

$$x_0^2 = x_1^2 a \mod b.$$

If $x_1 \not\equiv 0 \mod b$, then

$$a \equiv (\frac{x_0}{x_1})^2 \mod b$$

contradicts the assumption on a. Hence $b|x_1$ and $b|x_0$. The equation $Nr(x) = 0$ again implies that $x_2^2 \equiv ax_3^2 \mod b$, which in turn implies that $b|x_2, x_3$. This contradicts the assumption that x_0, x_1, x_2, x_3 have no common divisor and proves the proposition.

For example, if $b = 5$, $a = 2$ or 3, then $\mathcal{H}(a, b)$ is a division algebra. Certainly, we can give infinitely many examples of such division algebras $\mathcal{H}(a, b)$.

Now arithmetic Fuchsian groups are constructed as follows. Take a division algebra $\mathcal{H}(a, b)$ with a, b integers, $a > 0$, which is clearly defined over $\mathbb{Q}$. Define

$$\Gamma = \{x = x_0 + ix_1 + jx_2 + kx_3 \mid x_0, x_1, x_2, x_3 \in \mathbb{Z}, Nr(x) = 1\},$$

the norm 1 subgroup of the order

$$\mathcal{O} = \{x = x_0 + ix_1 + jx_2 + kx_3 \mid x_0, x_1, x_2, x_3 \in \mathbb{Z}\}$$

in $\mathcal{H}(a, b)$. Since $a > 0$, a is the square of a real number, and hence

$$\mathcal{H}(a, b) \otimes \mathbb{R} = M_{2\times 2}(\mathbb{R}).$$

In other words, $\mathcal{H}(a, b)$ gives a rational structure on $M_{2\times 2}(\mathbb{R})$ different from the standard one $M_{2\times 2}(\mathbb{Q})$. Since

$$\det \begin{pmatrix} x_0 + x_1\sqrt{a} & x_2 + x_3\sqrt{a} \\ b(x_2 - x_3\sqrt{a}) & x_0 - x_1\sqrt{a} \end{pmatrix} = Nr(h),$$

the image $\rho(\Gamma)$ under the embedding ρ in Equation (7) belongs to $SL(2, \mathbb{R})$.

Proposition 4.7 *If $\mathcal{H}(a, b)$ is a division algebra as above, then $\Gamma \cong \rho(\Gamma)$ is a discrete subgroup of $SL(2, \mathbb{R})$ with compact quotient $\Gamma\backslash SL(2, \mathbb{R})$.*

The discreteness is basically clear since Γ consists of integral elements. The compactness of the quotient is more complicated. It can either be proved directly (see [GGP, pp. 117-119] and [Ka, Theorem 5.4.1]) or follows from a general criterion given in the next section.

If $\mathcal{H}(a, b)$ is not a division algebra, then $\mathcal{H}(a, b) \cong M_{2\times 2}(\mathbb{Q})$, and the construction leads to arithmetic subgroups of $SL(2, \mathbb{R})$ commensurable with $SL(2, \mathbb{Z})$.

Arithmetic Fuchsian groups Γ can also be characterized in terms of the field generated by the trace of elements in Γ. See [Ka] for details.

Hilbert modular groups

Let F be a real quadratic field, $F = \mathbb{Q}(\sqrt{d})$, d is a square free positive integer. Then F has two real embeddings and no complex embedding. The group $SL(2, \mathbb{C})$ is defined over $\mathbb{Q}$ and hence also over F. The group $Res_{F/Q}SL(2)$ is defined over $\mathbb{Q}$ and of $\mathbb{Q}$-rank 1, and

$$Res_{F/Q}SL(2, \mathbb{R}) = SL(2, \mathbb{R}) \times SL(2, \mathbb{R}).$$

The arithmetic group $\Gamma = SL(2, \mathcal{O}_F)$ embeds into $SL(2, \mathbb{R}) \times SL(2, \mathbb{R})$ as a discrete subgroup, called the (principal) Hilbert modular group. It acts on the product $\mathbf{H} \times \mathbf{H}$ properly and the quotient $\Gamma\backslash\mathbf{H} \times \mathbf{H}$ has finite volume, called the Hilbert modular surface associated with F. The geometry of the Hilbert modular surface is closely related to the properties of the field F. For example, the number of ends of $\Gamma\backslash\mathbf{H}\times\mathbf{H}$ is equal to the class number of F (see [Fr, 3.5]). The Hilbert modular surface $\Gamma\backslash\mathbf{H} \times \mathbf{H}$ is probably the second most studied locally symmetric space after $SL(2,\mathbb{Z})\backslash\mathbf{H}$ considered earlier. Hilbert suggested to use it to understand real quadratic extensions of $\mathbb{Q}$. The $\mathbb{Q}$-rank of $Res_{F/Q}SL(2,\mathbb{R})$ is equal to 1, but the $\mathbb{R}$-rank is equal to 2 and hence strictly greater than 1. Hence the ends of $\Gamma\backslash\mathbf{H} \times \mathbf{H}$ are of $\mathbb{Q}$-rank one, topologically given by cylinders. More generally, we can consider a totally real number field F of degree d over $\mathbb{Q}$, i.e., F admits no complex embedding, $s = d$, $t = 0$. Then $SL(2,\mathcal{O}_F)$ is a discrete subgroup of $SL(2,\mathbb{R})^d$ and defines the Hilbert modular variety $\Gamma\backslash\mathbf{H}^d$. For more details and many questions about the Hilbert modular surfaces and varieties, see [Fr] [Ga] [Ge].

Bianchi group

From the point of view of group theory, the Bianchi group is a close analogue of the Hilbert modular group. Let $F = \mathbb{Q}(\sqrt{-d})$ be an imaginary quadratic field, where d is a positive square-free integer. Then $Res_{F/\mathbb{Q}}SL(2)$ is defined over $\mathbb{Q}$ of $\mathbb{Q}$-rank 1 and

$$Res_{F/\mathbb{Q}}(\mathbb{R}) = SL(2,\mathbb{C}).$$

The arithmetic subgroup $SL(2,\mathcal{O}_F)$ is a discrete subgroup of $SL(2,\mathbb{C})$ and called the Bianchi group. The symmetric space $X = G/K$ for $G = SL(2,\mathbb{C})$ is the real hyperbolic space of dimensional 3, i.e., the simply connected Riemannian manifold with constant curvature equal to -1, which can be realized as

$$\mathbf{H}_3 = \{(x,y,t) \mid x,y \in \mathbb{R}, t > 0\}, \quad ds^2 = \frac{dx^2 + dy^2 + dt^2}{t^2}.$$

The quotient $SL(2,\mathcal{O}_F)\backslash\mathbf{H}^3$ is a typical noncompact arithmetic 3-dimensional hyperbolic manifold of finite volume and has been extensively studied in topology (see [EGM] and [MR]). There are also co-compact arithmetic subgroups of $SL(2,\mathbb{C})$ constructed via quaternion algebras over F (see [MR]).

Picard modular groups

The Hilbert modular groups are associated with real quadratic fields and motivated by problems in number theory. On the other hand, the Picard modular groups are associated with imaginary quadratic fields and were originally motivated by the problems about differential equations with regular singularities in two variables (see [Ho1]); they are also very important examples of varieties for the Langlands program (see [La]). In comparison with the Bianchi group, the algebraic group is $SU(2,1)$ rather than $SL(2)$.

Let $\langle , \rangle$ be the Hermitian form on $\mathbb{C}^3$ defined by

$$\langle z, w\rangle = \bar{z}_1 w_1 + \bar{z}_2 w_2 - \bar{z}_3 w_3, \quad z = (z_1, z_2, z_3), \quad w = (w_1, w_2, w_3).$$

Let $SU(2,1)$ be the associated special unitary group

$$SU(2,1) = \{g \in SL(3, \mathbb{C}) \mid \langle gz, gw \rangle = \langle z, w \rangle\}.$$

Clearly $SU(2,1)$ is defined over $\mathbb{Q}$ and hence also defined over any imaginary quadratic field $F = \mathbb{Q}(\sqrt{-d})$, where d is a positive square free positive integer. $Res_{F/\mathbb{Q}} SU(2,1)$ is defined over $\mathbb{Q}$ and of $\mathbb{Q}$-rank 1, and

$$Res_{F/\mathbb{Q}} SU(2,1)(\mathbb{R}) = SU(2,1; \mathbb{C}),$$

which is often denoted by $SU(2,1)$ as above. The arithmetic subgroup $SU(2,1; \mathcal{O}_F)$ is a discrete subgroup of $SU(2,1)$ and called the Picard modular group associated with F. The symmetric space $X = G/K$ for $G = SU(2,1)$ is the unit ball in $\mathbb{C}^2$,

$$SU(2,1)/S(U(2) \times U(1)) \cong B^2_{\mathbb{C}} = \{(z_1, z_2) \in \mathbb{C}^2, |z_1|^2 + |z_2|^2 < 1\},$$

and the quotient $SU(2,1; \mathcal{O}_F)\backslash B^2_{\mathbb{C}}$ is called the Picard modular surface associated with the field F.

A slightly more general way to define the Picard modular group is as follows. Let V be a 3-dimensional vector space over the imaginary quadratic field $F = \mathbb{Q}(\sqrt{-d})$. Let L be an $\mathcal{O}_F$-lattice in V, which gives an integral structure on V. Let $J : V \times V \to F$ be a non-degenerate Hermitian form on V which takes values in $\mathcal{O}_F$ on L, i.e.,

1. $J(u, v) = \overline{J(v, u)}$, where the map $u \mapsto \bar{u}$ is the complex conjugation of $F/\mathbb{Q}$,
2. $J(\alpha u, \beta v) = \bar{\alpha}\beta J(u, v), \quad \alpha, \beta \in \mathbb{C}.$

Assume that J has signature $(2,1)$. The standard Hermitian form $\langle , \rangle$ is such a form. The algebraic group $SU(J, V \otimes \mathbb{C})$,

$$SU(J, V \otimes \mathbb{C}) = \{g \in SL((V \otimes \mathbb{C}) \mid J(gu, gv) = J(u, v), \text{ for all } u, v \in V \otimes \mathbb{C}\},$$

is defined over F, and

$$Res_{F/\mathbb{Q}} SU(J, V \otimes \mathbb{C})(\mathbb{R}) \cong SU(2,1; \mathbb{C}).$$

Then the subgroup

$$\Gamma = \{g \in SU(J, V) \mid gL = L\}$$

is called the Picard modular group associated with the triple V, J, L. (Note that $SU(J, V) = Res_{F/\mathbb{Q}} SU(J, V \otimes \mathbb{C})(\mathbb{Q})$.) See [LR] for detailed discussions about arithmetic geometric properties of the Picard modular surfaces $\Gamma\backslash B^2_{\mathbb{C}}$. One reason for the importance of the Picard modular surfaces is that they are locally symmetric spaces of both $\mathbb{Q}$-rank and $\mathbb{R}$-rank one and hence are accessible to detailed studies.

5 Symmetric spaces and compactness criterion of arithmetic quotients

Arithmetic groups can be used to define a very important class of locally symmetric spaces. In this and the next sections, we will study the geometry of such locally symmetric spaces.

Let $\mathbf{G}$ be a connected linear algebraic group defined over $\mathbb{Q}$, and $\Gamma \subset \mathbf{G}(\mathbb{Q})$ an arithmetic subgroup as in the previous section. Let $G = \mathbf{G}(\mathbb{R})$ be the real locus of $\mathbf{G}$.

Proposition 5.1 *The real locus G is a Lie group with finitely many connected components.*

Proof. It is a known fact in algebraic geometry that the real locus of a connected variety defined over $\mathbb{Q}$ has only finitely many connected components. Since G is a real variety, it has some smooth point and hence is smooth everywhere. Therefore, G is a Lie group with finitely many connected components.

Proposition 5.2 *The arithmetic subgroup Γ is a discrete subgroup of G.*

Proof. Clearly $GL(n, \mathbb{Z})$ is a discrete subgroup of $GL(n, \mathbb{R})$. Since $\mathbf{G}(\mathbb{Z}) = G \cap GL(n, \mathbb{Z})$ and the topology of G is the subset topology of $GL(n, \mathbb{R})$, $\mathbf{G}(\mathbb{Z})$ is a discrete subgroup of $G = \mathbf{G}(\mathbb{R})$. Since Γ is commensurable with $\mathbf{G}(\mathbb{Z})$, Γ is also a discrete subgroup of G.

For the rest of the section, we assume that $\mathbf{G}$ is a semisimple linear algebraic group. Then G is a semisimple Lie group. We assume that G is non-compact, equivalently the $\mathbb{R}$-rank of $\mathbf{G}$ is positive. Let K be a maximal compact subgroup of G. Then $X = G/K$ is a symmetric space of noncompact type when endowed with the Riemannian metric induced from the Killing form on the Lie algebra $\mathfrak{g}$ of G. For completeness, we recall several facts about symmetric spaces. For more details, see [Ji1] [Bo1] [He].

Let X be a symmetric space, and $Is(X)$ the group of isometries of X as in §2. Let $G = Is^0(X)$ the identity component. Fix a basepoint x_0 in X and let K be the stabilizer of x_0 in G. Then K is a compact subgroup, $X = G/K$, and the metric on X is G-invariant. On the other hand, given a pair of G, K, where $K \subset G$ is a compact subgroup, G/K always admits a G-invariant metric. In general, G/K is not a symmetric space.

Definition 5.3 Let G be a Lie group with finitely many connected components and K a closed subgroup. The pair $(G,\ K)$ is called a symmetric pair if there exists an involutive automorphism σ of G such that $G^{\sigma,0} \subset K \subset G^{\sigma}$. If, in addition, the image $Ad_G(K)$ under the map $Ad_G : G \to GL(\mathfrak{g})$ is compact, (G, K) is called a Riemannian symmetric pair.

For the pair G, K coming from a symmetric space X, the geodesic symmetry s_{x_0} defines an involution σ on $Is(X)$ and hence on G by

$$\sigma(g) = s_{x_0} g s_{x_0}.$$

Symmetric spaces X are classified into three types: the flat type, compact type and non-compact type, which basically correspond to the three cases that the sectional curvature of X is identically zero, non-negative, and non-positive. We are mainly concerned with symmetric spaces of noncompact type. For example, the upper half plane $\mathbf{H}$, the product $\mathbf{H} \times \mathbf{H}$, and the complex unit ball $B^2_{\mathbb{C}}$ in $\mathbb{C}^2$ are symmetric spaces of noncompact type.

It is also known that any semisimple noncompact Lie group G with finitely many connected components admits an involutive automorphism σ, unique up to conjugacy by G, such that the fixed point set G^σ is a maximal compact subgroup. Hence there is a unique symmetric space of noncompact type associated with G. Therefore, for any connected linear semisimple algebraic group $\mathbf{G}$ defined over $\mathbb{Q}$, there is a unique symmetric space $X = G/K$ of noncompact type.

There are two important classes of symmetric spaces. The first consists of Hermitian symmetric spaces of noncompact type, which can be realized as bounded symmetric domains. For example, $\mathbf{H}$ can be realized as the unit disc D in $\mathbb{C}$,

$$D = \{z \in \mathbb{C} \mid |z| < 1\},$$

and $\mathbf{H} \times \mathbf{H}$ as the bidisc $D \times D$ in $\mathbb{C}^2$. The second class consists of the so-called linear symmetric spaces.

The symmetric space $GL(n, \mathbb{R})/O(n)$ can be identified with the space of positive definite quadratic forms $\mathcal{P}_n$ on $\mathbb{R}^n$. Let $\mathcal{S}_n$ be the real vector space of $n \times n$ symmetric matrices. Then $\mathcal{P}_n$ is an open convex cone in $\mathcal{S}_n$, and the action of $GL(n, \mathbb{R})$ on $\mathcal{P}_n$ is the restriction of a linear action of $GL(n, \mathbb{R})$ on $\mathcal{S}_n$, where $GL(n, \mathbb{R})$ acts on $\mathcal{S}_n$ by

$$g \cdot A = g A {}^t g,$$

for $g \in GL(n, \mathbb{R}), A \in \mathcal{S}_n$. In fact, with respect to the inner product

$$\langle A, B \rangle = trAB, \quad A, B \in \mathcal{S}_n,$$

$\mathcal{P}_n$ is a *self-dual cone* in the sense that

$$\mathcal{P}_n = \{A \in \mathcal{S}_n \mid \langle A, B \rangle > 0, \text{ for all } \ B \in \mathcal{P}_n\}.$$

Definition 5.4 *A self-dual cone Ω in $\mathbb{R}^n$ with respect to a suitable inner product is called a symmetric cone if its automorphism group*

$$G(\Omega) = \{g \in GL(n, \mathbb{R}) \mid g\Omega = \Omega\}$$

acts transitively on Ω.

A symmetric cone Ω is a symmetric space of the form $G(\Omega)/K$, where K is a maximal compact subgroup of $G(\Omega)$. The homothety section of Ω is also a symmetric space. Both symmetric spaces are called linear symmetric spaces. In the example of $\mathcal{P}_n$, its homothety section is $SL(n,\mathbb{R})/SO(n)$, the space of positive definite symmetric matrices of determinant 1. The reason why they are called linear symmetric spaces is that the action of G and hence of a discrete subgroup Γ is linear. This can be exploited in constructing fundamental domains of arithmetic groups acting on linear symmetric spaces.

Since Γ is a discrete subgroup and K a compact subgroup of G, Γ acts properly on $X = G/K$, i.e., for any compact subset $C \subset X$, the set

$$\{\gamma \in \Gamma \mid gC \cap C \neq \emptyset\}$$

is finite.

If Γ is torsion free, then Γ acts freely on X, and the quotient $\Gamma\backslash X$ is a smooth manifold, a smooth locally symmetric space.

Proposition 5.5 *Any arithmetic subgroup $\Gamma \subset \mathbf{G}(\mathbb{Q}) \cap GL(n,\mathbb{Z})$ admits a torsion free subgroup Γ' of finite index.*

Proof. First we note that the order of torsion elements in Γ is uniformly bounded. In fact, for any torsion element $\gamma \in \Gamma$, the eigenvalues λ_i of γ are roots of unity of degree less than or equal to n, and hence there exists an integer $f(n)$ such that the order of λ_i is less than or equal to $f(n)$. Since Γ has n eigenvalues as a $n \times n$ matrix, the order of γ is less than or equal to $f(n)^n$. Choose any prime $p > f(n)^n$, and define

$$\Gamma_p = \{\gamma \in \Gamma \mid \gamma \equiv I \mod p\}.$$

Clearly Γ_p is of finite index in Γ. We claim that Γ_p is torsion free. If not, there exists an element $A \in \Gamma_p$, $A \neq Id$, $A^k = Id$. We can assume that $k < f(n)^n$. By the choice of p, we have $k < p$. Write $A = I + p^s B$, where $B \neq 0$, $s \geq 1$, $p \not| B$. Then

$$A^k = I + kp^s B \mod p^{2s} \not\equiv Id \mod p^{2s}.$$

This contradicts $A^k = Id$, and hence proves the claim.

A slightly stronger result also holds in Proposition 5.9 below (see [Bo2, §17]) and follows from similar arguments.

Definition 5.6 *An element of $GL(n,\mathbb{C})$ is neat if the subgroup of $\mathbb{C}$ generated by its eigenvalues is torsion free. An arithmetic subgroup Γ is called neat if every element of Γ is neat.*

Lemma 5.7 *If Γ is neat, then Γ is torsion free.*

Proof. For any element $\gamma \in \Gamma$, if $\gamma^k = 1$, then the eigenvalues of γ are roots of unity. Since γ is neat by assumption, all the eigenvalues of γ are equal to 1, and hence γ is unipotent. Then $\gamma^k = 1$ again implies that $\gamma = 1$.

Proposition 5.8 *Let $f : GL(n, \mathbb{C}) \to GL(m, \mathbb{C})$ be a morphism. If $\Gamma \subset GL(n, \mathbb{C})$ is neat, then $f(\Gamma)$ is also neat.*

This proposition [Bo2, Corollary 17.3] says that the neatness property is functorial. This implies that when passing to subgroups and quotient groups, neat arithmetic subgroups induce neat, and hence torsion free, subgroups.

Proposition 5.9 *Every arithmetic subgroup γ admits a neat subgroup of finite index.*

The first natural question about the geometry of $\Gamma \backslash X$ is whether it is compact or not. The answer is given in the following result.

Theorem 5.10 *Let $\mathbf{G}$ be a connected semisimple linear algebraic group defined over $\mathbb{Q}$, and $\Gamma \subset \mathbf{G}(\mathbb{Q})$ an arithmetic subgroup. Then the following conditions are equivalent:*

1. *The locally symmetric space $\Gamma \backslash X$ is compact.*
2. *$\mathbf{G}(\mathbb{Q})$ does not contain any nontrivial unipotent element.*
3. *The $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to 0.*

Since any two arithmetic subgroups Γ, Γ' of $\mathbf{G}(\mathbb{Q})$ are commensurable, $\Gamma \backslash X$ is compact if and only if $\Gamma' \backslash X$ is compact. Hence, whether $\Gamma \backslash X$ is compact or not depends only on the rational structure of $\mathbf{G}$, or equivalently $\mathbf{G}(\mathbb{Q})$. The condition that $\mathbf{G}(\mathbb{Q})$ does not contain any nontrivial unipotent element is equivalent to that Γ does not contain any nontrivial unipotent element.

Consider the example $\mathbf{G} = SL(2, \mathbb{C})$ and $\Gamma = SL(2, \mathbb{Z})$. It was known in Corollary 2.8 that $SL(2, \mathbb{Z}) \backslash \mathbf{H}$ is noncompact. In this case $SL(2, \mathbb{Q})$ contains nontrivial unipotent elements $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$, where $b \in \mathbb{Q}^\times$; and the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to 1. Another example is an arithmetic Fuchsian group Γ associated with a division quaternion algebra $\mathcal{H}(a, b)$. In this case, $\mathbf{G}(\mathbb{R}) = SL(2, \mathbb{R})$, and hence is semisimple. If Γ contains a nontrivial unipotent γ, then $\gamma - Id \neq 0$ but $Nr(\gamma - Id) = \det(\gamma - Id) = 0$, where $\gamma - Id$ is realized as a matrix in Equation (7). This contradicts the assumption that $\mathcal{H}(a, b)$ is a division algebra. Then the theorem implies Proposition 4.7 that $\Gamma \backslash \mathbf{H}$ is compact.

A slightly more general compactness criterion holds for reductive groups.

Proposition 5.11 *Let $\mathbf{G}$ be a connected reductive algebraic group defined over $\mathbb{Q}$, $\Gamma \subset \mathbf{G}(\mathbb{Q})$ an arithmetic subgroup. Then $\Gamma \backslash \mathbf{G}(\mathbb{R})$ is compact if and only if $\mathbf{G}$ does not admit a nontrivial character over $\mathbb{Q}$, and $\mathbf{G}(\mathbb{Q})$ does not contain any nontrivial unipotent element.*

The rest of this section is devoted to the proof of Theorem 5.10. We first show that (1) implies (2), and (2) is equivalent to (3).

Lemma 5.12 *If $\Gamma\backslash G$ is compact, then for any representation $\pi : \mathbf{G} \to GL(m, \mathbb{C})$ defined over $\mathbb{Q}$ and any $v \in \mathbb{Q}^m$, $v\pi(G)$ is a closed subset in $\mathbb{R}^m$, where the vectors in $\mathbb{Q}^m$ are row vectors.*

Proof. By the proof of Proposition 4.2, there exists a subgroup Γ' of Γ of finite index such that $\pi(\Gamma') \subset GL(m, \mathbb{Z})$, and hence $\mathbb{Z}^n\pi(\Gamma)$ is a closed, discrete subset. Since v is rational, $v\pi(\Gamma)$ is also a discrete, closed subset. By assumption, $\Gamma\backslash G$ is compact, hence there exists a compact subset $\omega \subset G$ such that $G = \Gamma\omega$. Then $v\pi(G) = (v\pi(\Gamma)) \cdot \omega$ is closed.

If (1) does not imply (2), then there is a nontrivial unipotent element u in $\mathbf{G}(\mathbb{Q})$. A theorem of Jacobson-Morosow [Ja, Lemma 7, p. 98] implies that there exists a morphism $\sigma : SL(2) \to \mathbf{G}$ defined over $\mathbb{Q}$ such that

$$\sigma(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}) = u.$$

Since the $SL(2, \mathbb{R})$-conjugacy class of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not closed,

$$\begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} t^{-1} & 0 \\ 0 & t \end{pmatrix} = \begin{pmatrix} 1 & t^2 \\ 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \text{ as } t \to 0,$$

the G-conjugacy class of u contains Id and hence is not closed either. Now consider the representation of $\mathbf{G}$ on $M_{n\times n}(\mathbb{C})$ via the adjoint map

$$\mathbf{G} \to GL(n^2, \mathbb{C}), \quad v \cdot g = gvg^{-1}.$$

This representation is defined over $\mathbb{Q}$, and hence the above lemma implies that the orbit $v \cdot G$, which is the G-conjugacy class of v, is closed. This contradiction shows that (1) implies (2).

We use the result of Jacobson-Morosow again to show that (3) implies (2). If not, let u be a nontrivial unipotent in $\mathbf{G}(\mathbb{Q})$. Let $\sigma : SL(2) \to \mathbf{G}$ be the morphism defined over $\mathbb{Q}$ associated with u as above. Then the $\mathbb{Q}$-split torus $\{\begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \mid t \neq 0\}$ of $SL(2)$ is mapped by σ to a $\mathbb{Q}$-split torus in $\mathbf{G}$, and hence the $\mathbb{Q}$-rank of $\mathbf{G}$ is positive.

To show that (2) implies (3), let $\mathbf{T}$ be a maximal $\mathbb{Q}$-split torus in $\mathbf{G}$. The associated root system of $\mathbf{G}$ with respect to $\mathbf{T}$ shows that $\mathbf{G}$ has non-trivial unipotent subgroups defined over $\mathbb{Q}$, which contains nontrivial rational unipotent element, and hence $\mathbf{G}(\mathbb{Q})$ contains nontrivial unipotent element.

The proof that (2) implies (1) is more complicated. We will first present a proof of a special case when Γ is assumed to be a lattice to indicate that existence of unipotent elements is closely related to noncompactness. Then we give the general proof.

Proposition 5.13 *Let G be a locally compact group with countable neighborhood basis and $\Gamma \subset G$ a lattice, i.e., a discrete subgroup of finite covolume with* *respect to any Haar measure. Let x_n be a sequence in G and $\pi : G \to \Gamma\backslash G$ the projection. Then $\pi(x_n)$ has no convergent sequence in $\Gamma\backslash G$ if and only if there exists a sequence $\gamma_n \in \Gamma$ such that $\gamma_n \neq e$, $x_n\gamma_n x_n^{-1} \to e$ as $n \to +\infty$.*

Proof. Let μ denote a Haar measure on G and the quotient measure on $\Gamma\backslash G$. Let B_n be an increasing family of compact subsets such that $\cup_{n=1}^{\infty} B_n = \Gamma\backslash G$, i.e., B_n is an exhausting family. Since $\mu(\Gamma\backslash G) < +\infty$,

$$\mu(\Gamma\backslash G - \pi(B_n)) \to 0.$$

Let V_n be a fundamental system of compact neighborhoods of e in G with

$$\mu(V_n) > \mu(\Gamma\backslash G - \pi(B_n)).$$

Then $V_n^{-1}V_n B_n$ and hence $\pi(V_n^{-1}V_n B_n)$ is also compact.

Suppose that $\pi(x_n)$ has no convergent subsequence. Then for $m \gg 1$,

$$\pi(x_m) \notin \pi(V_n^{-1}V_n B_n).$$

By multiplying over V_n^{-1}, we get that for $m \gg 1$,

$$\pi(V_n x_m) \cap \pi(V_n B_n) = \emptyset.$$

Since

$$\mu(V_n x_m) = \mu(V_n) > \mu(\Gamma\backslash G - \pi(B_n)) \geq \mu(\Gamma\backslash G - \pi(V_n B_n)),$$

it follows that for $m \gg 1$, there exists $\gamma_m \in \Gamma$, $\gamma_m \neq e$, and $v, v' \in V_n$ such that

$$v x_m \gamma_m = v' x_m,$$

and hence $x_m\gamma_m x_m^{-1} = v^{-1}v' \in V_n^{-1}V_n$ converges to e.

On the other hand, let $x_n \in G$ be any sequence such that there exists $\gamma_n \in \Gamma$, $\gamma_n \neq e$ satisfying that $x_n\gamma_n x_n^{-1} \to e$. We claim that $\pi(x_n)$ has no convergent subsequence. If not, by passing to a subsequence if necessary, we assume that $\pi(x_n) \to \pi(x_\infty)$ for some $x_\infty \in G$. By replacing x_n by a suitable Γ-translate, we can assume that $x_n \to x_\infty$. Then the convergence $x_n\gamma_n x_n^{-1} \to e$ implies that $\gamma_n \to x_n^{-1} e x_n = e$. This contradicts the assumption that Γ is a discrete subgroup.

Proposition 5.14 *Let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be an arithmetic subgroup. If $\Gamma\backslash G$ has finite volume and $\mathbf{G}(\mathbb{Q})$ has no nontrivial unipotent element, then $\Gamma\backslash G$ is compact.*

Proof. By taking a subgroup of finite index if necessary, we can assume that $\Gamma \subset GL(n, \mathbb{Z})$. If $\Gamma\backslash G$ is not compact, then there exists a sequence $x_n \in G$ such that no subsequence of $\pi(x_n)$ is bounded in $\Gamma\backslash G$. By the previous proposition, there exists $\gamma_n \in \Gamma$ such that $\gamma_n \neq e$, $x_n\gamma_n x_n^{-1} \to e$. The characteristic polynomial P_n of $x_n\gamma_n x_n^{-1}$ is the same as the characteristic polynomial of γ_n and hence has integral coefficients. The convergence $x_n \gamma x_n^{-1} \to e$ implies that when $n \gg 1$, $P_n = 1$, which in turn implies that all the eigenvalues of γ_n are equal to 1, i.e., γ_n is unipotent. This contradicts the assumption, and the proposition is proved.

By reduction theory, which will be discussed in the next section, when $\mathbf{G}$ is semisimple, any arithmetic subgroup is co-finite in $\mathbf{G}$. Hence the implication from (2) to (1) in Theorem 5.10 follows from Proposition 5.14. But we can give a direct proof without using the reduction theory. The idea is as follows:

1. embed $\Gamma\backslash G$ into an ambient space $\mathcal{L}$,

2. get some compactness criterion for subsets in $\mathcal{L}$ and check that the image of $\Gamma\backslash G$ satisfies this criterion.

Let $\mathcal{L}$ be the space of lattices in $\mathbb{R}^n$. Since $GL(n,\mathbb{R})$ acts transitively on $\mathcal{L}$ with the stabilizer of the standard lattice $\mathbb{Z}^n$ equal to $GL(n,\mathbb{Z})$, $\mathcal{L}$ can be identified with $GL(n,\mathbb{Z})\backslash GL(n,\mathbb{R})$. For the proof of Theorem 5.10, we can assume that $\Gamma = \mathbf{G}(\mathbb{Z})$. Then $\Gamma\backslash G$ can be identified with the G-orbit of the standard lattice $\mathbb{Z}^n$ in $\mathcal{L}$, and hence

$$\Gamma\backslash G \hookrightarrow GL(n,\mathbb{Z})\backslash GL(n,\mathbb{R}) \cong \mathcal{L}.$$

Lemma 5.15 *The image $\Gamma\backslash G$ is a closed subset of $\mathcal{L}$.*

Proof. We need to show that $GL(n,\mathbb{Z})G$ is a closed subset of $GL(n,\mathbb{R})$. Since $\mathbf{G}$ is an algebraic subgroup of $GL(n)$ defined over $\mathbb{Q}$, it is a known fact in algebraic group theory (see Lemma 5.17 below) that there exists a representation $\rho : GL(n,\mathbb{C}) \to GL(m,\mathbb{C})$ over $\mathbb{Q}$ and a vector $v \in \mathbb{Q}^m$ such that its stabilizer in $GL(n)$ is equal to $\mathbf{G}$. Since there exists a lattice Λ in $\mathbb{R}^m$ containing v which is invariant under $GL(n,\mathbb{Z})$, it follows that $\rho(GL(n,\mathbb{Z}))v$ is a closed subset in $\mathbb{R}^m$. Consider the map $GL(n,\mathbb{R}) \to \mathbb{R}^m$, $g \mapsto \rho(g)v$. This is a continuous map, and the inverse image of $\rho(GL(n,\mathbb{Z})v)$ is equal to $GL(n,\mathbb{Z})G$. This implies that $GL(n,\mathbb{Z})G$ is a closed subset of $GL(n,\mathbb{R})$.

The compactness criterion of subsets of $\mathcal{L}$ is given by the following Mahler criterion (see [Bo2, Corollary 1.9]), which easily follows from the determination of the fundamental set of $SL(n,\mathbb{Z})$ in $SL(n,\mathbb{R})$ in Proposition 6.3 below.

Proposition 5.16 *A subset $M \subset GL(n,\mathbb{R})$ has a bounded (i.e. relative compact) image in the quotient $GL(n,\mathbb{Z})\backslash GL(n,\mathbb{R})$ if and only if the following two conditions are satisfied: (1) $\{|\det g| \mid g \in M\}$ is bounded, (2) for any v_j in the lattice $\mathbb{Z}^n$, and $g_j \in M$, if $g_j v_j \to 0$, then $v_j = 0$ for $j \gg 1$.*

We apply this criterion to $M = G$. Assume that $\mathbf{G}$ satisfies condition (2) in Theorem 5.10. Since the center of $\mathbf{G}$ is finite, we can assume for simplicity that $\mathbf{G}$ has trivial center. Then the adjoint representation $Ad : G \to GL(\mathfrak{g})$ is faithful and defined over $\mathbb{Q}$. Choose an integral structure on (or a lattice Λ in)

$\mathfrak{g}$ and assume that $Ad(\Gamma) \subset GL(\mathfrak{g}, \mathbb{Z})$. It suffices to show that the image of $\Gamma\backslash G$ in $GL(\mathfrak{g}, \mathbb{Z})\backslash GL(\mathfrak{g})$ is compact.

Since $\mathbf{G}$ is semisimple, $|\det g| = 1$ for $g \in G$, and hence the first condition is satisfied. We need to check the second condition. For any $v \in \mathfrak{g}$, let $P(v)$ be the characteristic polynomial of $ad(v)$:

$$P(v) = \det(ad(v) - \lambda) = (-\lambda)^n + \sum_{i=1}^{n-1} P_i(v)\lambda^i,$$

where $P_i(v)$ are polynomials with integer coefficients. Since $\mathbf{G}(\mathbb{Q})$ does not contain any nontrivial unipotent element, $\mathfrak{g}_{\mathbb{Q}}$ does not contain any nontrivial nilpotent element (recall that the exponential map is given by polynomials on nilpotent elements). This implies that for $v \in \Lambda \setminus \{0\}$, $P(v)$ is a polynomial in λ with integer coefficients and

$$P(v) \neq (-\lambda)^n. \tag{8}$$

If $g_j \in G$, $v_j \in \Lambda \setminus \{0\}$ satisfy $Ad(g_j)v_j \to 0$, then

$$P(v_j) = P(Ad(g_j)(v_j)) \to (-\lambda)^n.$$

Since $P(v_j)$ has integral coefficients, it follows that

$$P(v_j) = (-\lambda)^n$$

for $j \gg 1$. This contradicts Equation (8), and hence the second condition is also satisfied. This shows $\Gamma\backslash G$ is compact and completes the proof of Theorem 5.10.

For completeness, we outline a proof of the following fact from algebraic group theory used in the above proof.

Lemma 5.17 *Let $\mathbf{H}$ be a linear algebraic group defined over $\mathbb{Q}$, $\mathbf{G} \subset \mathbf{H}$ a reductive algebraic subgroup defined over $\mathbb{Q}$. Then there exists a finite dimensional vector space W over $\mathbb{Q}$, a representation of $\mathbf{H}$ on W defined over $\mathbb{Q}$ and a vector $w \in W_{\mathbb{Q}}$ such that the stabilizer of w is equal to $\mathbf{G}$.*

Proof. Let $\mathbb{C}[\mathbf{H}]$ be the algebra of regular functions on $\mathbf{H}$. Then $\mathbf{G}$ acts on it. Since $\mathbf{G}$ is reductive, the subalgebra of $\mathbf{G}$-invariant functions $\mathbb{C}[\mathbf{H}]^{\mathbf{G}}$ is finitely generated over $\mathbb{Q}$. Let $w_1, \cdots, w_m$ be generators over $\mathbb{Q}$. Now each w_i is contained in a finite dimensional $\mathbf{H}$-invariant subspace W_i defined over $\mathbb{Q}$. Let $W = W_1 \oplus \cdots \oplus W_m$. Let $w = (w_1, \cdots, w_m)$ be the vector formed from the generators. It is clear that $\mathbf{G}$ fixes w. On the other hand, if an element h of $\mathbf{H}$ fixes w, $hw = w$, then

$$hw_i = w_i,$$

and hence they take the same value at e,

$$w_i(h) = hw_i(e) = w_i(e).$$

Since $\mathbb{C}[\mathbf{H}]^{\mathbf{G}}$ separate $\mathbf{G}$-stable closed subsets, it follows that $h \in \mathbf{G}$.

6 Reduction theories for arithmetic groups

In the previous section, we showed that an arithmetic group Γ of a semisimple linear algebraic group $\mathbf{G}$ is a co-compact in $\mathbf{G}(\mathbb{R})$ if and only if the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to zero. In other words, when the $\mathbb{Q}$-rank of $\mathbf{G}$ is positive, $\Gamma\backslash X$ is non-compact. In this section, we describe a nice fundamental set for Γ in X and use it to show that $\Gamma\backslash X$ has finite volume, i.e., Γ is a lattice.

As mentioned in §1, the problem of finding a fundamental domain or set is called reduction theory, which plays an important role in understanding the geometry and topology of and analysis on $\Gamma\backslash X$.

There are three approaches to the reduction theory:

1. Find a fundamental set for general arithmetic groups. This is the classical reduction theory due to Borel & Harish-Chandra, and refined by Borel [Bo2].

2. Find a fundamental domain, or a fundamental set with the covering multiplicity equal to 1. This is called the precise reduction theory and was motivated by applications to the Arthur-Selberg trace formula [Sap2] [OW].

3. For linear symmetric spaces, use the geometry of numbers to find fundamental domains in terms of polyhedral cones. Besides its applications to locally linear symmetric space, it also plays an important role in compactifications of Hermitian locally symmetric spaces, or rather Shimura varieties [AMRT].

To generalize the fundamental domain of $SL(2,\mathbb{Z})$ in $\mathbf{H}$ discussed in §2, we need to understand the coordinates x, y of the upper half plane $\mathbf{H}$ in terms of group structure.

In $SL(2,\mathbb{C})$, the subgroup

$$\mathbf{P} = \{\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \mid a \in \mathbb{C}^\times, b \in \mathbb{C}\}$$

is a parabolic subgroup defined over $\mathbb{Q}$. Let $P = \mathbf{P}(\mathbb{R})$ be its real locus. Then P is the stabilizer of $i\infty$ in $SL(2,\mathbb{R})$ under the fractional linear action on $\mathbb{C}\cup\{i\infty\}$. For $\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \in P$, and $z \in \mathbf{H}$,

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} z = a^2 z + ab.$$

The parabolic subgroup P contains two subgroups

$$N_P = \{\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \mid b \in \mathbb{R}\}, \quad A_P = \{\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \mid a \in \mathbb{R}^\times\}.$$

For any $z \in \mathbf{H}$, the orbit through z of N_P is a horizontal line and the x-coordinate is related to b. On the other hand, unless $z = iy$, the A_P-orbit of z

is not a vertical coordinate line. To overcome this difficulty, we introduce the Langlands decomposition of P,

$$P = N_P A_P M_P \cong N_P \times A_P \times M_P, \quad M_P = \{\pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\}.$$

Clearly the P-orbit through i is equal to $\mathbf{H}$. Then the Langlands decomposition of P induces an identification

$$N_P \times A_P \cong \mathbf{H}, \quad (n, a) \to nai. \tag{9}$$

Under this identification, for any $a \in A_P$, $N \times \{a\}$ is a horizontal line, a horocircle at $i\infty$; while for any $n \in N_P$, $\{n\} \times A_P$ is a vertical line, a geodesic converging to $i\infty$. Hence, the decomposition in Equation (9) gives the x, y-coordinates and the horospherical decomposition of $\mathbf{H}$ with respect to the point $i\infty$.

In this horospherical decomposition $\mathbf{H} = N_P \times A_P$, the region

$$\mathcal{S}_t = \{x + iy \mid -\frac{1}{2} \le x \le \frac{1}{2}, y > t\}$$

can be expressed as

$$\mathcal{S}_t = U \times A_{P,t},$$

where

$$U = \{\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \mid -\frac{1}{2} \le b \le \frac{1}{2}\}, \quad A_{P,t} = \{\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \mid a > t^{1/2}\}.$$

When $t < \frac{\sqrt{3}}{2}$, $U \times A_{P,t}$ covers the fundamental domain Ω of $SL(2, \mathbb{Z})$ in §2, and hence is mapped surjectively onto $\Gamma\backslash\mathbf{H}$. This fact was used to show that $\Gamma\backslash\mathbf{H}$ has finite area in Corollary 2.8. Since the map from $U \times A_{P,t}$ to $\Gamma\backslash\mathbf{H}$ is not injective on its interior, it is not a fundamental domain. Rather it is a fundamental set defined in the following sense.

Definition 6.1 *Let $X = G/K$ be a symmetric space and Γ an arithmetic subgroup of $\mathbf{G}(\mathbb{Q})$. Then a subset $\mathcal{S}$ of X is called a fundamental set if*

1. *$\Gamma\mathcal{S} = X$.*
2. *For any $g \in \mathbf{G}(\mathbb{Q})$, the set $\{\gamma \in \Gamma \mid g\mathcal{S} \cap \gamma\mathcal{S} \ne \emptyset\}$ is finite.*

To define the fundamental sets of a fixed arithmetic group Γ, we can replace (2) above by a weaker condition:

(2') The set $\{\gamma \in \Gamma \mid \mathcal{S} \cap \gamma\mathcal{S} \ne \emptyset\}$ is finite.

But we need condition (2) to relate fundamental sets of different arithmetic subgroups and different algebraic groups, for example, to derive fundamental sets of general Γ from the special case $\mathbf{G} = SL(n), \Gamma = SL(n, \mathbb{Z})$. This condition is called the Siegel (finiteness) property and plays an important role in defining compactifications of $\Gamma\backslash X$ and showing that the topologies of the compactifications are Hausdorff.

In general, for any bounded set U in N_P and any $t > 0$, a subset in $\mathbf{H}$ of the form $U \times A_{P,t}$ is called a Siegel set associated with the parabolic subgroup P.

For the rest of this section, we discuss the following topics:

1. The Langlands decomposition of rational parabolic subgroups, and the associated horophserical decomposition of X.
2. Siegel sets.
3. Fundamental sets for $\Gamma = SL(n, \mathbb{Z})$ in $SL(n, \mathbb{R})/SO(n)$.
4. Fundamental sets for general pairs Γ, G.
5. Exact fundamental sets for general pairs Γ, G (the precise reduction theory).
6. Reduction theory for linear symmetric spaces.

Let $\mathbf{P}$ be a rational parabolic subgroup of $\mathbf{G}$, i.e., a parabolic subgroup of $\mathbf{G}$ defined over $\mathbb{Q}$. Let $\mathbf{N_P}$ be the unipotent radical of $\mathbf{P}$, i.e., the largest normal unipotent subgroup in $\mathbf{P}$. Then the quotient $\mathbf{L_P} = \mathbf{N_P}\backslash\mathbf{P}$ is reductive and called the Levi quotient. We want to write $\mathbf{P} = \mathbf{N_P} \times \mathbf{L_P}$ and use it to get the Langlands decomposition of $P = \mathbf{P}(\mathbb{R})$, the real locus. For this purpose, we need to lift $\mathbf{L_P}$ into a subgroup of $\mathbf{P}$. Suppose $i : \mathbf{L_P} \to \mathbf{P}$ is a lift, then for any $p \in \mathbf{P}$, $p\, i(\mathbf{L_P})p^{-1}$ is also a lift of $\mathbf{L_P}$. So there is no unique lift unless we impose some condition.

Recall that $G = \mathbf{G}(\mathbb{R})$, $K \subset G$ a maximal compact subgroup, and $X = G/K$. The subgroup K determines a basepoint $x_0 = K$ and a Cartan involution θ on G whose fixed point set is equal to K. Recall that the Cartan involution on $\mathfrak{g}$ determined by K gives a decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$, where $\mathfrak{k}$ is the Lie algebra of K, and $\mathfrak{p}$ can be identified with the tangent space of X at x_0. It is known that θ extends to an involution on $\mathbf{G}$.

Proposition 6.2 *For the basepoint x_0, there exists a unique lift $i_{x_0} : \mathbf{L_P} \to \mathbf{P}$ such that the image $i_{x_0}(\mathbf{L_P})$ is stable under the Cartan involution θ.*

The Levi quotient $\mathbf{L_P}$ is a reductive algebraic group defined over $\mathbb{Q}$. Let $\mathbf{S_P}$ be the maximal $\mathbb{Q}$-split torus in the center $\mathbf{Z}(\mathbf{L_P})$. Let $A_\mathbf{P}$ be the identity component of the real locus of $\mathbf{S_P}(\mathbb{R})$, $M_\mathbf{P}$ the complement in $\mathbf{L_P}(\mathbb{R})$, i.e.,

$$\mathbf{L_P}(\mathbb{R}) = A_\mathbf{P} M_\mathbf{P} \cong A_\mathbf{P} \times M_\mathbf{P}.$$

The subgroup $A_\mathbf{P}$ is called the split component of P, and $\dim_\mathbb{R} A_\mathbf{P}$ is called the $\mathbb{Q}$-rank of $\mathbf{P}$. Under the lift i_{x_0}, we get subgroups $A_{\mathbf{P},x_0} = i_{x_0}(A_\mathbf{P})$, $M_{\mathbf{P},x_0} = i_{x_0}(M_\mathbf{P})$ in P and the Langlands decomposition of P with respect to the basepoint:

$$P = N_\mathbf{P} \times A_{\mathbf{P},x_0} \times M_{\mathbf{P},x_0},$$

where $N_{\mathbf{P}} = \mathbf{N}_{\mathbf{P}}(\mathbb{R})$. In the following, we will fix this basepoint $x_0 = K$ and hence drop the subscript x_0 from $A_{\mathbf{P},x_0}$, $M_{\mathbf{P},x_0}$.

It is a known fact that $G = PK$, and hence P acts transitively on $X = G/K$. Let $K_{\mathbf{P}} = M_{\mathbf{P}} \cap K$, and

$$X_{\mathbf{P}} = M_{\mathbf{P}}/K,$$

called the boundary symmetric space associated with $\mathbf{P}$. Since $M_{\mathbf{P}}$ is reductive, $X_{\mathbf{P}}$ is in general product of a symmetric space of noncompact type and an Euclidean space. For example, this is the case with the Hilbert modular group. Then the Langlands decomposition of P induces the horospherical decomposition of X:

$$X \cong N_{\mathbf{P}} \times A_{\mathbf{P}} \times X_{\mathbf{P}},$$

where a point $(n, a, mK_{\mathbf{P}})$ is mapped to $namK \in X$.

In the example of $X = \mathbf{H}$, and $\mathbf{P}$ the subgroup of upper triangular matrices, $X_{\mathbf{P}}$ consists of one point, and the horospherical decomposition is reduced to the earlier one

$$\mathbf{H} \cong N_P \times A_P.$$

The split component $A_{\mathbf{P}}$ acts on the Lie algebra of $N_{\mathbf{P}}$ by the adjoint action. Denote the roots of this action by $\Phi(P, A_{\mathbf{P}})$. For any $t > 0$, define a cone in $A_{\mathbf{P}}$ by

$$A_{P,t} = \{a \in A_{\mathbf{P}} \mid a^{\alpha} > t, \quad a \in \Phi(P, A_{\mathbf{P}})\}.$$

When $t = 1$, $A_{P,t}$ is the image under the exponential map of the positive Weyl chamber in the Lie algebra $\mathfrak{a}_{\mathbf{P}}$ of $A_{\mathbf{P}}$. Let $U \subset N_{\mathbf{P}}$, $V \subset X_{\mathbf{P}}$ be bounded subsets. Then the subset

$$\mathcal{S}_{P,U,V,t} = \mathcal{S}_{U,V,t} = U \times A_{\mathbf{P},t} \times V \subset X$$

is called the Siegel set associated with P. As seen above, for $X = \mathbf{H}$ and $\mathbf{P}$ the parabolic subgroup of upper triangular matrices, the Siegel sets are given by a part of vertical strips. The goal of the reduction theory for $\mathbf{G}$ is to construct fundamental sets of arithmetic groups in terms of Siegel sets.

We will illustrate the above concepts and constructions using the example of $\mathbf{G} = SL(n)$, which is defined over $\mathbb{Q}$ as explained earlier. The subgroup $\mathbf{P}$ of upper triangular matrices is a minimal parabolic subgroup defined over $\mathbb{Q}$:

$$\mathbf{P} = \{ \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ & & & a_{nn} \end{pmatrix} \mid a_{ij} \in \mathbb{C}\}.$$

Then the unipotent radical $\mathbf{N}_{\mathbf{P}}$ is given by

$$\mathbf{N}_{\mathbf{P}} = \{ \begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ & 1 & \cdots & a_{2n} \\ & & \cdots & \\ & & & 1 \end{pmatrix} \mid a_{ij} \in \mathbb{C}\}.$$

Take $K = SO(n)$. Then the Levi quotient $\mathbf{L_P}$ is lifted to the subgroup

$$\mathbf{L}_{\mathbf{P},x_0} = \{diag(a_{11}, \cdots, a_{nn}) \mid a_{11}, \cdots, a_{nn} \in \mathbb{C}, a_{11} \cdots a_{nn} = 1\}.$$

The split component $A_{\mathbf{P}}$ is given by

$$A_{\mathbf{P}} = \{diag(a_{11}, \cdots, a_{nn}) \mid a_{11}, \cdots, a_{nn} > 0, a_{11} \cdots a_{nn} = 1\},$$

and

$$M_{\mathbf{P}} = \{diag(a_{11}, \cdots, a_{nn}) \mid a_{ii} = \pm 1, a_{11} \cdots a_{nn} = 1\}.$$

The boundary symmetric space $X_{\mathbf{P}}$ consists of one point. The set of roots $\Phi(P, A_{\mathbf{P}})$ is given by

$$\Phi(P, A_{\mathbf{P}}) = \{a_{ii}/a_{jj} \mid i < j\},$$

and

$$A_{\mathbf{P},t} = \{diag(a_{11}, \cdots, a_{nn}) \in A_{\mathbf{P}} \mid a_{11}/a_{22} > t, \cdots, a_{n-1,n-1}/a_{nn} > t\}.$$

For the above minimal parabolic subgroup P, there are only finitely many parabolic subgroups containing P, called the standard parabolic subgroups, which are given by upper triangular block matrices. For example, for each $k, 1 \leq k \leq n-1$, the subgroup

$$\mathbf{Q} = \{\begin{pmatrix} A & C \\ 0 & B \end{pmatrix} \mid A \in GL(k), B \in GL(n-k), \det A \det B = 1, C \in M_{k\times n-k}(\mathbb{C})\}$$

is a maximal parabolic subgroup defined over $\mathbb{Q}$. Its associated subgroups are

$$N_{\mathbf{Q}} = \{\begin{pmatrix} I_k & C \\ 0 & I_{n-k} \end{pmatrix} \mid C \in M_{k\times n-k}(\mathbb{C})\},$$

$$A_{\mathbf{Q}} = \{\begin{pmatrix} a_1 I_k & 0 \\ 0 & a_2 I_{n-k} \end{pmatrix} \mid a_1, a_2 > 0, a_1 a_2 = 1\},$$

$$M_{\mathbf{Q}} = \{\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \mid |\det A| = 1, |\det B| = 1, \det A \det B = 1\}.$$

The boundary symmetric space

$$X_{\mathbf{Q}} = SL(k, \mathbb{R})/SO(k) \times SL(n-k)/SO(n-k),$$

a reducible symmetric space.

There are several approaches to the reduction theory of $\mathbf{G}, \Gamma$. The standard one described in the books [PR] [Bo2] (the fundament sets of the second kind) proceeds in two steps:

1. Use the idea of the reduction for $SL(2), \Gamma = SL(2, \mathbb{Z})$ to describe a fundamental set for $\mathbf{G} = SL(n), \Gamma = SL(n, \mathbb{Z})$.

2. Embed $\mathbf{G}$ suitably into $SL(n)$ and use the reduction for $SL(n, \mathbb{Z})$.

This approach is not direct since it is not easy to see that the subsets of X obtained via intersection with the fundamental sets for $SL(n, \mathbb{Z})$ in $SL(n, \mathbb{R})/SO(n)$ are Sigel sets of Γ in X. There is another more intrinsic approach, called the fundamental sets of the third kind in [Bo2, §16], based on minimum of suitable functions, as in the proof of Proposition 2.5. It was pointed out on [Bo2, p. 108], this approach depended on the fundamental set of the second type obtained in the first approach to show that there are only finitely many Γ-conjugacy classes of rational parabolic subgroups, which is crucial to this second approach. The finiteness of Γ-conjugacy classes of rational parabolic subgroups is equivalent to that for a minimal rational parabolic subgroup $\mathbf{P}$, the double cosets $\Gamma\backslash\mathbf{G}(\mathbb{Q})/\mathbf{P}(\mathbb{Q})$ is finite. By [Go, Theorems 5, 6], this finiteness result holds if Γ is a congruence subgroup, for example $\mathbf{G}(Z)$, hence for any subgroup Γ' of finite index of $\mathbf{G}(\mathbb{Z})$, $\Gamma'\backslash\mathbf{G}(\mathbb{Q})/\mathbf{P}(\mathbb{Q})$ is also finite. Since any arithmetic subgroup Γ contains a subgroup Γ' of finite index which is also contained in $\mathbf{G}(\mathbb{Z})$ as a subgroup of finite index, it follows that

$$|\Gamma\backslash\mathbf{G}(\mathbb{Q})/\mathbf{P}(\mathbb{Q})| \leq |\Gamma'\backslash\mathbf{G}(\mathbb{Q})/\mathbf{P}(\mathbb{Q})| < +\infty.$$

This simple observation shows that the second approach is actually independent of the first one.

We will start with the first approach and briefly outline the second approach. Let $\mathbf{P}$ be the minimal rational parabolic subgroup of $SL(n)$ above. Then $X_{\mathbf{P}}$ consists of one point, and

$$X = N_{\mathbf{P}} \times A_{\mathbf{P}}$$

is the associated horospherical decomposition. For any $u > 0$, let

$$U = N_{\mathbf{P},u} = \{ \begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ & 1 & a_{23} \cdots & a_{2n} \\ & & \cdots & \\ & & & 1 \end{pmatrix} \mid |a_{ij}| \leq u \}.$$

Proposition 6.3 *For $\Gamma = SL(n, \mathbb{Z})$, $\mathbf{G} = SL(n)$, when $u \geq \frac{1}{2}$, $t < \frac{\sqrt{3}}{2}$, the Siegel set $U \times A_{P,t}$ is mapped surjectively onto $\Gamma\backslash X$.*

Before proving this proposition, we note that it does not imply that $U \times A_{P,t}$ is a fundamental set for $SL(n, \mathbb{Z})$. We also need the Siegel finiteness result.

The idea of the proof of this proposition is as follows. Clearly $N_{\mathbf{P}} = N_{\mathbf{P}}(\mathbb{Z})N_{\mathbf{P},\frac{1}{2}}$, and hence translates of $N_{\mathbf{P},\frac{1}{2}} \times A_{\mathbf{P}}$ under $SL(n, \mathbb{Z})$ covers X. The problem is to replace the group $A_{\mathbf{P}}$ by a cone $A_{\mathbf{P},t}$. In the case of $\mathbf{G} = SL(2)$, $A_{\mathbf{P}}$ is a line, and the cone $A_{\mathbf{P},t}$ is a half line. For $n \geq 3$, $A_{\mathbf{P},t}$ is a smaller portion of $A_{\mathbf{P}}$. This problem is solved by choosing in each Γ-orbit Γx an element γx with the "largest" $A_{\mathbf{P}}$-component, which is equivalent to the maximum imaginary part in the case of $X = \mathbf{H}$ in Proposition 2.5. Since elements in $A_{\mathbf{P}}$ are vectors, we use representations of G with suitable weights to quantify the maximum values.

Proof of Proposition 6.3.

Let $e_1 = \begin{pmatrix} 1 \\ 0 \\ \cdot \\ o \end{pmatrix}, \cdots, e_n = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ 1 \end{pmatrix}$ be the standard orthonormal basis of $\mathbb{R}^n$.

Define a function

$$\phi : G = SL(n, \mathbb{R}) \to \mathbb{R}^+, \quad g \mapsto ||g^{-1}e_1||.$$

(Note that in the identity representation of $SL(n, \mathbb{R})$, e_1 is a highest weight vector with respect to the order determined by the positive roots $a_{11}/a_{22}, \cdots, a_{n-1,n-1}/a_{nn}$.)

For $a = diag(a_{11}, \cdots, a_{nn}) \in A_{\mathbf{P}}$,

$$\phi(a) = ||diag(a_{11}^{-1}, \cdots, a_{nn}^{-1})e_1|| = a_1^{-1}.$$

For $n \in N_{\mathbf{P}}$, $ne_1 = e_1$ and hence $\phi(n) = 1$. For $k \in K = SO(n)$,

$$\phi(gk) = ||k^{-1}g^{-1}e_1|| = ||g^{-1}e_1|| = \phi(g).$$

Hence ϕ descends to a function on $X = G/K$. The above discussions show that in the decomposition $g = (n, a, m) \in N_{\mathbf{P}} \times A_{\mathbf{P}} \times M_{\mathbf{P}}K \cong G$,

$$\phi(nam) = \phi(a) = a_1^{-1}.$$

This implies that the function ϕ only depends on the $A_{\mathbf{P}}$-component in the horopsherical decomposition of X. Note that the inversion g^{-1} in the definition of ϕ is used crucially here.

The idea of the proof is to show that on one each orbit Γg in G, ϕ achieves its minimum at a point γg such that $\gamma g x_0 \in N_{\mathbf{P},u} \times A_{\mathbf{P},t}$, for any $u \geq \frac{1}{2}$, $t < \frac{\sqrt{3}}{2}$.

Lemma 6.4 *The function ϕ does achieve a minimum value at some point in any orbit Γg.*

Proof. For any $g \in G$,

$$(\Gamma g)^{-1}e_0 = g^{-1}\Gamma e_0 = g^{-1}(\mathbb{Z}^n).$$

Clearly, the norm function $\|\cdot\|$ achieves a minimum on the lattice $g^{-1}(\mathbb{Z}^n) \subset \mathbb{R}^n$.

Lemma 6.5 *If ϕ takes a minimal value at a point $x_1 = \gamma x$ on the orbit Γx, then there exists another point $x_2 \in \Gamma x$ with*

$$\phi(x_2) = \phi(x_1)$$

such that the horospherical coordinates n, a of x_2, $x_2 = (n, a) \in N_{\mathbf{P}} \times A_{\mathbf{P}}$, satisfy

1. $n \in N_{\mathbf{P}, \frac{1}{2}}$.

2. $a_{11}/a_{22} \geq \frac{\sqrt{3}}{2}$.

Proof. Since ϕ is invariant under the left action of $SL(n,\mathbb{Z}) \cap N_{\mathbf{P}}$, we can find a point $x_1 \in \Gamma x$ such that its $N_{\mathbf{P}}$-component $u \in N_{\mathbf{P},\frac{1}{2}}$. Next we use the minimality to show that the second condition is satisfied. Take

$$Z = \begin{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & 0 \\ 0 & I_{n-2} \end{pmatrix} \in \Gamma = SL(n,\mathbb{Z}).$$

By assumption, $\phi(Zx_1) \geq \phi(x_1)$. Write $x_1 = gx_0$. Then

$$\phi(Zx_1) = \phi(Zg) = ||(Zg)^{-1}e_1|| = ||g^{-1}Z^{-1}e_1||.$$

Since $N_{\mathbf{P}}A_{\mathbf{P}}$ acts transitively on X, we can take

$$g^{-1} = diag(a_{11}^{-1}, \cdots, a_{nn}^{-1}) \begin{pmatrix} 1 & u_{12} & \cdots & u_{1n} \\ & 1 & \cdots & u_{2n} \\ & & \cdots & . \\ & & & 1 \end{pmatrix},$$

where $|u_{ij}| \leq \frac{1}{2}$. Since $Z^{-1}E_1 = e_2$,

$$(Zg)^{-1}e_1 = g^{-1}e_2 = a_{11}^{-1}u_{12}e_1 + a_{22}^{-1}e_2 \tag{10}$$

$$\phi(Zg)^2 = ||a_{11}^{-1}u_{12}e_1 + a_{22}^{-1}e_2||^2 = a_{11}^{-2}u_{12}^2 + a_{22}^{-2} \leq \frac{1}{4}a_{11}^{-2} + a_{22}^{-2}, \tag{11}$$

and hence

$$a_{11}^{-2} = \phi(g)^2 = \phi(Zg) \leq \frac{1}{4}a_{11}^{-2} + a_{22}^{-2},$$

and

$$\frac{3}{4}a_{11}^{-2} \leq a_{22}^{-2},$$

which implies that

$$a_{11}/a_{22} \geq \frac{\sqrt{3}}{2}.$$

Once we have this lemma, we can use induction to show

$$a_{22}/a_{33}, \cdots, a_{n-1,n-1}/a_{nn} \geq \frac{\sqrt{3}}{2}$$

and that every orbit Γx contains at least one point in the Siegel set $N_{\mathbf{P},u} \times A_{\mathbf{P},t}$, or we can use other functions

$$||g^{-1}e_i||, \quad i = 2, \cdots, n-1$$

and their minimum values to get the desired bounds on $a_{ii}/a_{i+1,i+1}$ In a certain sense, a point in $N_{\mathbf{P},u} \times A_{\mathbf{P},t}$ is a simultaneous minimum point of these functions

$||g^{-1}e_i||$, $i = 1, \cdots, n-1$, and hence the $A_{\mathbf{P}}$-component of g is "maximal". The proof of Proposition 6.3 is complete.

To show that $N_{\mathbf{P},u} \times A_{\mathbf{P},t}$ is a fundamental set for $SL(n,\mathbb{Z})$, we need to check the second condition in Definition 6.1 as well.

Proposition 6.6 *For any Siegel set $\mathcal{S}_{u,t} = N_{\mathbf{P},u} \times A_{\mathbf{P},t}$ and any two elements $g_1, g_2 \in \mathbf{G}(\mathbb{Q})$, the set*

$$\{\gamma \in SL(n,\mathbb{Z}) \mid g_1\mathcal{S}_{u,t} \cap \gamma g_2 \mathcal{S}_{u,t} \neq \emptyset\}$$

is finite.

The proof of this proposition uses a result of Harish-Chandra and is complicated. The basic idea is this. Since $\Gamma = SL(n,\mathbb{Z})$ is discrete, it suffices to prove that this set is bounded. Since the $N_{\mathbf{P}}$-component of the elements in the Siegel sets is uniformly bounded, the $N_{\mathbf{P}}$-component of elements in this set is also bounded, and hence it suffices to bound the $A_{\mathbf{P}}$-part of elements of this set. For this purpose, one first shows that the functions $||g^{-1}e_i||$ used in the proof of Proposition 6.3 satisfy some multiplicative bounds on the Siegel sets, and reduce the problem to getting a lower bound on the norm of the $A_{\mathbf{P}}$-component. Even though the matrices in Γ have integral entries, getting a lower bound is not easy. The reason is that if the $A_{\mathbf{P}}$-component has integral entries too, then the lower bound is obvious. But the $A_{\mathbf{P}}$-component is defined using the Langlands decomposition, and this decomposition does not preserve the integral (or rational) property of the group elements. On the other hand, the Bruhat decomposition preserves the integral structure of elements, which in the case of $SL(n)$ is obtained by separating out different parts of the matrices, hence we can get a lower bound on the $A_{\mathbf{P}}$-component in the Bruhat decomposition. Then the problem is to relate these two decomposition. For details see [Bo2] and [PR]. The same argument works for general $\mathbf{G}, \Gamma$.

The Siegel finiteness property allows us to construct a fundamental set for any arithmetic subgroup $\Gamma \subset SL(n,\mathbb{Q})$ using the fundamental set constructed already for $SL(n,\mathbb{Z})$.

Proposition 6.7 *For any arithmetic two arithmetic subgroups Γ, Γ'. If Ω is a fundamental set for Γ, then*

$$\Omega' = \cup_{\xi \in \Gamma\cap\Gamma' \backslash \Gamma} \xi\Omega$$

is a fundamental set for Γ'.

Proof. Write $\Gamma = \Gamma \cap \Gamma'\xi_1 \cup \cdots \cup \Gamma \cap \Gamma'\xi_m$, where $m = [\Gamma, \Gamma \cap \Gamma']$. Since $X = \Gamma\Omega$,

$$X = (\Gamma \cap \Gamma')(\xi_1\Omega \cup \cdots \cup \xi_m\Omega) \subseteq \Gamma'(\xi_1\Omega \cup \cdots \cup \xi_m\Omega) \subseteq X.$$

The Siegel finiteness property of $\xi\Omega_1 \cup \cdots \cup \xi_m\Omega$ follows easily from the Siegel finiteness property of Ω.

We give two immediate applications of the existence of fundamental sets here. More will be given in the next section.

Proposition 6.8 *For any arithmetic subgroup $\Gamma \subset SL(n, \mathbb{Q})$, the volume of $\Gamma \backslash SL(n, \mathbb{R})/SO(n)$ is finite.*

Proof. By explicit computation, we can show that the volume of the Siegel set $N_{\mathbf{P},u} \times A_{\mathbf{P},t}$ is finite. Then the finiteness of the volume follows from the previous proposition.

Proposition 6.9 *Any arithmetic subgroup Γ of $SL(n, \mathbb{Q})$ is finitely generated.*

Proof. Since Siegel sets can be taken to be open subsets, there exists an open subset Ω of Γ in X such that $\Gamma\Omega = X$, and the set

$$\Pi = \{\gamma \in \Gamma \mid \gamma\Omega \cap \Omega \neq \emptyset\}$$

is finite. We claim that Π generates Γ. Let Γ_0 be the subgroup of Γ generated by Π. If $\Gamma_0 \neq \Gamma$. Then

$$X = \Gamma_0\Omega \cup (\Gamma \setminus \Gamma_0)\Omega.$$

We claim that this is a disjoint union. If not, then there exist some $\delta \in \Gamma_0$, $\gamma \in \Gamma - \Gamma_0$,

$$\delta\Omega \cap \gamma\Omega \neq \emptyset, \quad \gamma^{-1}\delta\Omega \cap \Omega \neq \emptyset.$$

Hence $\gamma^{-1}\delta \in \Pi$, and $\gamma \in \Gamma_0$. This is a contradiction.

Next we use the reduction for $\mathbf{G} = SL(n), \Gamma = SL(n, \mathbb{Z})$ to get a fundamental set for general semisimple linear algebraic groups $\mathbf{G}$ and arithmetic subgroups Γ.

For simplicity, assume $\mathbf{G} \subset SL(n)$, $\Gamma = \mathbf{G}(\mathbb{Z}) = \mathbf{G}(\mathbb{Q}) \cap SL(n, \mathbb{Z})$, $K = G \cap SO(n)$. Then

$$X \hookrightarrow SL(n, \mathbb{R})/SO(n), \quad \Gamma\backslash G/K \hookrightarrow GL(n, \mathbb{Z})\backslash GL(n, \mathbb{R})/SO(n).$$

Let Ω be a fundamental set for $SL(n, \mathbb{Z})$ in $SL(n, \mathbb{R})/SO(n)$. A naive question is whether $X \cap \Omega$ is a fundamental set for Γ.

For any $x \in X \subset SL(n, \mathbb{R})/SO(n)$, there exist $\gamma \in SL(n, \mathbb{Z})$ and $y \in \Omega$ such that $x = \gamma y$. If $\gamma \in G$ as well, i.e., $\gamma \in \Gamma$, then $y \in \Omega \cap X$; if $y \in X$, i.e., $y \in X \cap \Omega$, then $\gamma \in \Gamma$. In these cases, $x \in \Gamma(X \cap \Omega)$. But $x \in X$ alone does not imply either $\gamma \in G$ or $y \in X$. There could exist $\gamma \notin \Gamma$, $y \notin X$ such that $\gamma y \in X$. For this idea or some generalization to work, the point is to get some control on the components y, γ. Specifically, we need the following result.

Lemma 6.10 *Let Γ be a discrete subgroup in a Lie group H and Ω a subset in H such that $H = \Gamma\Omega$. Let Y be a topological H-space. Let $y_0 \in Y$ be a basepoint and $G = H_{y_0}$ be the stabilizer of y_0 in H. Assume that for a suitable $a \in H$, the intersection $\Gamma y_0 \cap \Omega a y_0$ is finite and equal to $\{b_1 y_0, \cdots, b_r y_0\}$, $b_i \in \Gamma$. Then*

$$G = (\Gamma \cap G)[(\cup_{i=1}^r b_i^{-1}\Omega a) \cap G].$$

Proof. Since $H = \Gamma\Omega$, $H = \Gamma\Omega a$. For any $g \in G$, write $g = \gamma\sigma a$, where $\gamma \in \Gamma$, $\sigma \in \Omega$. Then $gy_0 = y_0$ implies

$$\gamma\sigma y_0 = y_0.$$

Hence

$$\sigma a y_0 = \gamma^{-1} y_0 \in \gamma^{-1} y_0 \cap \Omega a y_0.$$

By assumption, there exists b_i such that

$$\gamma^{-1} y_0 = b_i y_0.$$

This implies that

$$\gamma b_i y_0 = y_0, \quad \gamma b_i \in \Gamma \cap G,$$

and hence

$$\gamma \in (\Gamma \cap G) b_i^{-1}.$$

(Note that this overcomes the difficulty in the comments above.) Therefore,

$$g \in (\Gamma \cap G) b_i^{-1} \sigma a \subset (\Gamma \cap G) b_i^{-1} \Omega a,$$

$$G \subseteq (\Gamma \cap G)[\cup_{i=1}^{r} b_i^{-1} \Omega a \cap G].$$

To apply this lemma to the present situation, the ambient group is $SL(n, R)$ (or $GL(n, \mathbb{R})$) and the stabilizer of a point y_0 should be G.

The following is a fact from algebraic group theory and a variant was recalled earlier in Lemma 5.18.

Proposition 6.11 *Let $\mathbf{G} \subset GL(n, \mathbb{C})$ be a linear algebraic group defined over $\mathbb{Q}$. Then there exists a representation $\rho : GL(n, \mathbb{C}) \to GL(V)$, where V is a vector space defined over $\mathbb{Q}$ and a vector $v \in V_{\mathbb{Q}}$ such that*

$$\mathbf{G} = \{g \in GL(n, \mathbb{C}) \mid \rho(g) v = v\}$$

and the orbit $\rho(GL(n, \mathbb{C}))v$ is Zariski closed in V.

Proposition 6.12 *Let $\rho : GL(n, \mathbb{C}) \to GL(V)$ be a representation as in the previous proposition and $L \subset V_{\mathbb{Q}}$ a lattice. If $w \in V_{\mathbb{R}}$ is a vector whose stabilizer*

$$\mathbf{G} = \{g \in GL(n, \mathbb{C}) \mid \rho(g) w = w\}$$

is a self-adjoint group (invariant under the conjugate transpose) and $\rho(GL(n, \mathbb{C}))$ is Zariski closed, then $\rho(\mathcal{S}_{u,t}) w \cap L$ is finite.

Note that for any $v \in V_{\mathbb{Q}}$, $\rho(SL(n, \mathbb{Z}))v$ is contained in a lattice. Hence the conclusion in the above proposition is almost the condition we need in the above lemma. See [PR, §4.3, Proposition 4.5].

Reductive groups are not necessarily self-adjoint.

Proposition 6.13 *For any reductive algebraic group* $\mathbf{G} \subset GL(n, \mathbb{C})$*, there exists* $a \in GL(n, \mathbb{R})$ *such that* $a\mathbf{G}a^{-1}$ *is self-adjoint.*

Let $v \in V_{\mathbb{Q}}$ and $w = av$. Then the stabilizer of w is equal to $a\mathbf{G}a^{-1}$. The above discussions give the following.

Proposition 6.14 *Let* $\mathbf{G} \subset GL(n, \mathbb{C})$ *be a reductive algebraic group defined over* $\mathbb{Q}$*, and* $\Gamma \subset \mathbf{G}(\mathbb{Q})$ *an arithmetic subgroup. Let* $\mathcal{S}_{u,t} = N_{\mathbf{P},u} \times A_{\mathbf{P},t}$ *a fundamental set for* $GL(n, \mathbb{Z})$*. Then there exist finitely many elements* $b_1, \cdots, b_r \in GL(n, \mathbb{Q})$ *and an element* $a \in G$ *such that*

$$\Gamma[(\cup_{i=1}^r b_i^{-1} \mathcal{S}_{u,t} a) \cap X] = X.$$

A natural problem is how to relate the set in the above proposition to Siegel sets in $\mathbf{G}$.

Proposition 6.15 *Let* $\mathbf{P}$ *be a minimal rational parabolic subgroup of a reductive algebraic group* $\mathbf{G} \subset GL(n, \mathbb{C})$ *defined over* $\mathbb{Q}$*,* $\Gamma \subset \mathbf{G}(\mathbb{Q})$ *an arithmetic subgroup. Then the double coset* $\Gamma\backslash\mathbf{G}(\mathbb{Q})/\mathbf{P}(\mathbb{Q})$ *is finite. Let* C *be a finite set of representatives of this double coset. Then there exists a Siegel set* $\mathcal{S}_{\mathbf{P},t}$ *such that*

$$(\cup_{i=1}^r b_i^{-1} \mathcal{S}_{u,t} a) \cap X \subset C\mathcal{S}_{\mathbf{P},t}.$$

Hence $C\mathcal{S}_{\mathbf{P},t}$ *is a fundamental set for* Γ *in* X*.*

The proof of the first statement is complicated (see [Bo2] [PR]). For the second statement, condition (1) in the definition of fundamental sets follows from the previous proposition, and condition (2) follows from Proposition 6.6.

Corollary 6.16 *Let* $\Gamma \subset \mathbf{G}(\mathbb{Q})$ *be an arithmetic subgroup of a semisimple algebraic group* $\mathbf{G}$*. Then* $\Gamma\backslash X$ *has finite volume.*

It follows from the fact that the volume of a Siegel set of X has finite volume.

The finite set C in Proposition 6.15 can be identified with the Γ-conjugacy classes of minimal rational parabolic subgroups, and Proposition 6.15 can be reformulated as follows. Let $\mathbf{P}_1, \cdots, \mathbf{P}_m$ be a set of representatives of Γ-conjugacy classes of minimal rational parabolic subgroups of $\mathbf{G}$. Then for each $\mathbf{P}_i$, there exists a Siegel set $\mathcal{S}_{\mathbf{P}_i,U_i,V_i,t_i}$ such that

$$X = \Gamma(\cup_{i=1}^m \mathcal{S}_{\mathbf{P}_i,U_i,V_i,t_i}).$$

When $\mathbf{G} = SL(2)$, every (proper) rational parabolic subgroup is minimal and $X = \mathbf{H}$. There is an one-to-one correspondence between the Γ-conjugacy classes of rational parabolic subgroups and the set of ends of the Riemann surface $\Gamma\backslash\mathbf{H}$, where each end is a cusp neighborhood. In this case, the boundary of the image in $\Gamma\backslash\mathbf{H}$ of each Siegel set is a horocircle. The heights of these horocircles are low enough so that the images of the Siegel sets can cover the whole space. On the other hand, when the heights of these horocircles are pushed up sufficiently high, the images of the Siegel sets become disjoint, and there exists a bounded set Ω_0 in $\Gamma\backslash\mathbf{H}$ such that $\Gamma\backslash\mathbf{H}$ admits a disjoint decomposition

$$\Gamma\backslash\mathbf{H} = \Omega_0 \cup \coprod_{i=1}^{m} \pi(\mathcal{S}_{\mathbf{P}_i,U_i,t_i}),$$

where $t_i \gg 0$, and $\pi : \mathbf{H} \to \Gamma\backslash\mathbf{H}$. Take open subsets $\tilde{\Omega}_0 \subset \mathbf{H}$, and $\Omega_i \subset \mathcal{S}_{\mathbf{P}_i,U_i,t_i}$ in $\mathbf{H}$ such that each of the following maps is injective and has dense images:

$$\pi : \tilde{\Omega}_0 \to \Omega_0, \quad \pi : \Omega_i \to \pi(\mathcal{S}_{\mathbf{P}_i,U_i,t_i}).$$

Then $\tilde{\Omega}_0 \cup \coprod_{i=1}^{r} \Omega_i$ is a fundamental domain for Γ in $\mathbf{H}$.

This modification can be generalized to get the precise reduction theory for general $\mathbf{G}, \Gamma$. For this purpose, we need to define Siegel sets slightly differently.

Let $\mathbf{P}_1, \cdots, \mathbf{P}_l$ be a set of representatives of Γ-conjugacy classes of all proper (not only minimal) rational parabolic subgroups. For each $\mathbf{P}_i$, and $T_i \in A_{\mathbf{P}_i}$, define

$$A_{\mathbf{P}_i,T_i} = \{a \in A_{\mathbf{P}_i} \mid a^\alpha > T_i^\alpha,\ \alpha \in \Phi(P, A_{\mathbf{P}})\},$$

and the corresponding Siegel set

$$\mathcal{S}_{\mathbf{P}_i,U_i,V_i,T_i} = U_i \times A_{\mathbf{P}_i,T_i} \times V_i.$$

Proposition 6.17 *With the above notation, there exist a bounded set Ω_0 and Siegel sets $\mathcal{S}_{\mathbf{P}_i,U_i,V_i,T_i}$ such that each is mapped injectively into $\Gamma\backslash X$ under $\pi : X \to \Gamma\backslash X$, and X admits a disjoint decomposition*

$$X = \Omega_0 \cup \coprod_{j=1}^{l} \pi(\mathcal{S}_{\mathbf{P}_i,U_i,V_i,t_i}).$$

Hence the union of the interior of Ω_0 and $\mathcal{S}_{\mathbf{P}_i,U_i,V_i,T_i}$ is a fundamental domain for Γ

This is called the precise reduction theory for Γ acting on X. The reason why we need all the rational parabolic subgroups can be seen in the decomposition of the split component $A_{\mathbf{P}}$ of minimal parabolic subgroup $\mathbf{P}$ of $\mathbf{G}$. For example, when $\mathbf{G} = SL(3)$ and $\mathbf{P}$ is the subgroup of upper triangular matrices, $A_{\mathbf{P}}$ has dimension 2. Let P_1, P_2 be the two maximal parabolic subgroups containing $\mathbf{P}$. Then the positive chamber $A_{\mathbf{P},1} = \exp \mathfrak{a}_{\mathbf{P}}^+$ can be decomposed into four pieces:

$$A_{\mathbf{P},1} = \Omega_0 \cup \omega_1 \times A_{\mathbf{P}_1,t_1} \cup \omega_2 \times A_{\mathbf{P}_2,t_2} \cup A_{\mathbf{P},T},$$

where T is determined by t_1, t_2, and ω_1, ω_2 are suitable bounded subsets, Ω_0 is a bounded set associated with $\mathbf{G}$.

Remark 6.18 Besides the motivation to find a fundamental domain, the precise reduction theory has applications to the Selberg trace formula. For detailed discussions about the precise reduction theory, see [OW] [Sap2]. We comment that the precise reduction theory as formulated here does not appear in [Sap2], and [Sap2] instead gives a Γ-equivariant tiling of X satisfying certain properties. It is easy to construct a fundamental domain from this equivariant tiling. For example, Γ acts cocompactly on the central tile, and a bounded fundamental domain for the Γ-action on this cental tile gives Ω_0 in Proposition 6.17, and the Siegel sets $\mathcal{S}_{\mathbf{P}_i,U_i,V_i,T_i}$ are constructed from tiles of the parabolic subgroups $\mathbf{P}_i$.

The second approach to the reduction theory for general $\mathbf{G}, \Gamma$ is given by the following result and is similar to Proposition 2.5 (see [Bo1, §16, Theorem 16.7]). The key point is the assumption that for a minimal rational parabolic subgroup $\mathbf{P}$, the double coset $\Gamma\backslash\mathbf{G}(\mathbb{Q})/\mathbf{P}(\mathbb{Q})$ is finite. As pointed out earlier, this follows from the result for congruence subgroups in [Go]. Hence, this approach gives an independent, more direct proof of the reduction theory.

Proposition 6.19 *Let $\mathbf{G}$ be a semisimple linear algebraic group defined over $\mathbb{Q}$, $\mathbf{P}$ a minimal rational parabolic subgroup. Let $\pi : \mathbf{G} \to GL(V)$ be an absolutely irreducible representation defined over $\mathbb{Q}$ such that there exists a vector $v \in V_{\mathbb{Q}}$ which is an eigenvector of $\pi(\mathbf{P})$. Let $||\cdot||$ be a K-invariant norm on V such that $\pi(A_{\mathbf{P}})$ are diagonal matrices with respect to an orthonormal basis. Define $\phi(g) = ||\pi(g^{-1})v||$. Let C be a set of representatives of the double coset $\Gamma\backslash\mathbf{G}(\mathbb{Q})/\mathbf{P}(\mathbb{Q})$, which is assumed to be finite. Then there exists a Siegel set $\mathcal{S}_{\mathbf{P},U,V,t}$ such that on any orbit $C\Gamma x$ in X, ϕ achieves its minimum value at some point in $C\Gamma x \cap \mathcal{S}_{\mathbf{P},U,V,t}$. Hence*

$$X = \Gamma C^{-1}\mathcal{S}_{\mathbf{P},U,V,t},$$

and $C^{-1}\mathcal{S}_{\mathbf{P},U,V,t}$ is a fundamental set for Γ in X.

Finally we describe the reduction theory for linear symmetric spaces through the example of $\mathbf{G} = GL(n,\mathbb{R})$. For the general theory, see [AMRT]. The goal is to get a fundamental domain using polyhedral cones.

Let $\mathcal{P}_n$ be the convex cone of positive definite $n \times n$-matrices. Recall that a rational polyhedral cone C in $\mathcal{P}_n$ is a cone spanned by a finite number of rational rays in the closure of $\mathcal{P}_n$, i.e., there exist finitely many rational positive semi-definite matrices $A_1, \cdots, A_r$ such that

$$C = \{\sum_{i=1}^{r} a_i A_i \mid a_i > 0\}.$$

Note that even though each A_i is semidefinite, if there are enough A_i and they are independent in a suitable sense, C is an open cone in $\mathcal{P}$. Each face of C is spanned by a proper subset of $A_1, \cdots, A_r$.

The reduction theory for $\mathcal{P}_n$ is reduced to the following problems:

1. Find a collection Σ of rational polyhedral cones in $\mathcal{P}_n$ which is disjoint, locally finite and $GL(n,\mathbb{Z})$-invariant.

2. Show that there are only finitely many $GL(n,\mathbb{Z})$-equivalence classes of rational polyhedral cones in the collection Σ. For any arithmetic subgroup Γ, let $C_1, \cdots, C_l$ be a set of representatives of the Γ-equivalence classes of the cones in Σ. Then the union $C_1 \cup \cdots \cup C_l$ (or rather its interior) is a fundamental domain for Γ in $\mathcal{P}_n$. (By taking a homothety section given by the vectors of norm 1 in $C_1 \cup \cdots \cup C_l$, we get a fundamental domain for $\Gamma \cap SL(n,\mathbb{Q})$ in $SL(n,\mathbb{R})/SO(n)$.)

There are several approaches to find the collection Σ of cones. We will discuss one. Each $A \in \mathcal{P}_n$ defines a positive definite quadratic form on $\mathbb{R}^n$. Let

$$m(A) = \min\{A(v,v) \mid v \in \mathbb{Z}^n - \{0\}\},$$

called the arithmetic minimum of A,

$$M(A) = \{v \in \mathbb{Z}^n - \{0\} \mid A(v,v) = m(A)\},$$

the set of minimum (integral) vectors. Let $\sqrt{A}$ be the positive matrix such that $(\sqrt{A})^2 = A$. Then $\sqrt{A}\mathbb{Z}$ is a lattice in $\mathbb{R}^n$, and the density of the sphere packing on the lattice $\sqrt{A}\mathbb{Z}^n$ is proportional to

$$\frac{m(A)^{n/2}}{|\det A|^{1/2}}.$$

In the problem of sphere packing, we want to maximize this density. The local maximum ponits are called *extreme forms*.

Definition 6.20 A form is called perfect if it is determined by $m(A)$ and $M(A)$, i.e.,

$$\{B \in \mathcal{P}_n \mid m(B) = m(A), M(B) = M(A)\} = \{A\}.$$

It is known that extreme forms are perfect. The space of lattices in $\mathbb{R}^n$ of covolume 1 can be identified with $SL(n,\mathbb{R})/SL(n,\mathbb{Z})$. Near the infinity of $SL(n,\mathbb{R})/SL(n,\mathbb{Z})$, the density of the sphere packing goes to 0. Hence extreme forms exist.

For each perfect form A, we can construct a rational polyhedral cone. In fact, let $M(A) = \{X_1, \cdots, X_s\}$. Then each $X_i\, {}^tX_i$ is a positive semidefinite matrix. Define a cone

$$C(A) = \{\sum_{i=1}^{s} a_i X_i\, {}^tX_i \mid a_i > 0\}.$$

Then $C(A)$ is contained in $\mathcal{P}_n$, and its open faces contained in $\mathcal{P}_n$ are also rational cones.

Proposition 6.21 *For any arithmetic subgroup of $SL(n,\mathbb{Z})$, the collection Σ of rational polyhedral cone $C(A)$ and its faces for perfect forms A gives a Γ-invariant disjoint decomposition of $\mathcal{P}_n$, and there are only finitely many Γ-equivalence classes in Σ.*

The Γ-invariance is clear since $SL(n, \mathbb{Z})$ maps a perfect form to another perfect form.

As mentioned earlier, this reduction theory of linear symmetric spaces in terms of polyhedral cones is important to the toroidal compactifications of Hermitian locally symmetric spaces, in particular, Shimura varieties. One way to see this connection is that for each symmetric cone, there is an associated tube domain, called a Siegel domain of the first kind, which can be realized as a bounded symmetric domain. Siegel domains of the second and third kinds are built up from the Siegel domains of the first kind. The polyhedral cones are needed to define torus embeddings which are used crucially in the toroidal compactifications in [AMRT].

7 Metric properties of locally symmetric spaces

In the rest of these lectures, we discuss applications of the reduction theory developed in the previous section. In this section, we study metric properties of $\Gamma\backslash X$ and related compactifications. More compactifications of $\Gamma\backslash X$ will be studied in the next section.

In the following, we assume that $\mathbf{G}$ is a semisimple linear algebraic group defined over $\mathbb{Q}$, $\Gamma \subset \mathbf{G}(\mathbb{Q})$ an arithmetic subgroup of $\mathbf{G}(\mathbb{Q})$, and $X = G/K$ the symmetric space of noncompact type of maximal compact subgroups of G.

A natural question is why we study metric properties of $\Gamma\backslash X$. One short answer is that $\Gamma\backslash X$ are important Riemannian manifolds, and we want to understand their common metric properties. Due to connections with group theories, many geometric properties can be understood well.

Another motivation comes from complex analysis. In complex analysis of one variable, there is a well-known result called the big Picard theorem, which says that near an essentially singular point of a meromorphic function, its values can miss at most three points of $\mathbb{C}P^1$. This result can be expressed in terms of metric properties of $\mathbb{C}P^1 \setminus \{0, 1, \infty\}$, which is of the form $\Gamma\backslash\mathbf{H}$. Let $D = \{z \in \mathbb{C} \mid |z| < 1\}$ be the unit disc, and $D^* = D \setminus \{0\}$ be the punctured disc. Then the big Picard theorem is equivalent to the statement that every holomorphic map $f : D^* \to \mathbb{C}P^1 \setminus \{0, 1, \infty\}$ can be extended to a holomorphic map $f : D \to \mathbb{C}P^1$. This is related to the fact that the Kobayashi pseudo-metric on $\mathbb{C}P^1 \setminus \{0, 1, \infty\}$ is a metric. Briefly, the Kobayashi metric is the maximal pseudo-metric such that (1) it coincides with the Poincare metric for the unit disc, (2) and is distance decreasing under holomorphic maps. (See [Kob] for details and the definition of Kobayashi pseudo-metric).

This result can be generalized to the following situation. Let Y be a complex space and M be a complex subspace of Y whose closure $\overline{M}$ is compact. The question is *whether every holomorphic map $f : D^* \to M$ extends to a holomorphic map $f : D \to Y$.*

In the above example, $Y = \mathbb{C}P^1$, and $M = \mathbb{C}P^1 \setminus \{0, 1, \infty\}$, and the answer is positive. Though the answer is negative in general, it holds under some conditions.

Proposition 7.1 *The answer to the above question is positive if the following two conditions hold:*

1. *M is a hyperbolic manifold, i.e., the Kobayashi pseudo-metric d_M of M is a metric.*

2. *For any two sequences $p_j, q_j \in M$ with $p_j \to p_\infty$, $q_j \to q_\infty$ in Y, and $d_M(p_j, q_j) \to 0$, then $p_\infty = q_\infty$.*

The conditions say that the compactification $\overline{M}$ of M in Y is small in a certain sense. If these conditions are satisfied, the embedding $M \hookrightarrow Y$ is called a *hyperbolic embedding with respect to the metric d_M*.

As mentioned earlier, many important spaces $\Gamma\backslash X$ arise as moduli spaces of certain varieties or structures in algebraic geometry and are often non-compact. A natural, important problem is to understand how these objects degenerate, or their moduli points go to the infinity of $\Gamma\backslash X$. This is equivalent to understanding whether holomorphic maps

$$f : D^* \to \Gamma\backslash X$$

extend and the senses in which they extend. For example, such extensions are important in the theory of variation of Hodge structures (see [GS]).

This raises two questions:

1. Construct compactifications $\overline{\Gamma\backslash X}$ of $\Gamma\backslash X$ if noncompact.

2. Understand metric properties of the compactifications, for example whether the embedding $\Gamma\backslash X \hookrightarrow \overline{\Gamma\backslash X}$ is hyperbolic with respect the invariant metric.

Let $\pi : X \to \Gamma\backslash X$ be the projection. Let d_X, $d_{\Gamma\backslash X}$ be the distance function of X and $\Gamma\backslash X$ respectively induced from the invariant metric. For any two points $p, q \in X$,

$$d_{\Gamma\backslash X}(\pi(p), \pi(q)) = \min\{d_X(p, \gamma q) \mid \gamma \in \Gamma\},$$

and hence

$$d_{\Gamma\backslash X}(\pi(p), \pi(q)) \leq d_X(p, q).$$

In general, it is difficult (or impossible) to bound $d_{\Gamma\backslash X}(\pi(p), \pi(q))$ from below in terms of (some functions) of $d_X(p, q)$, for example, when $q = \gamma p$, $\gamma \in \Gamma$, $\gamma \neq e$.

On the other hand, Siegel conjectured that when p, q belong to a Siegel set, then $d_{\Gamma\backslash X}(\pi(p), \pi(q))$ and $d_X(p, q)$ are comparable up to an additive constant. This conjecture was proved in [Ji2].

Proposition 7.2 *Let $\mathbf{P}$ be a rational parabolic subgroup of $\mathbf{G}$ and $\mathcal{S}_{\mathbf{P},t}$ an associated Siegel set. Then there exists a constant $C > 0$ such that for any $p, q \in \mathcal{S}_{\mathbf{P},t}$,*

$$d_X(p, q) - C \leq d_{\Gamma\backslash X}(\pi(p), \pi(q)) \leq d_X(p, q).$$

The next question concerns structures of geodesics in $\Gamma\backslash X$. The symmetric space X is simply connected and non-positively curved and is hence a so-called Hadamard manifold. Hadamard initiated the study of geodesics in such manifolds, and their structure plays an important role in understanding the geometry of manifolds (see [BGS] for details).

There are also other reasons to study geodesics. In fact, the spectral geometry studies relations between the geometry and spectral theory of Riemannian manifolds M. When M is compact, its natural geometric invariants are its volume and the lengths of closed geodesics, which form the length spectrum. On the other hand, the Laplacian operator Δ of M has discrete spectrum consisting of eigenvalues $\{\lambda_i\}$, $\Delta\varphi_i = \lambda_i\varphi_i$. In this case, the Weyl law on the growth of the counting function of the eigenvalues λ_i with multiplicity says that the leading term is determined by the volume of M and its dimension and hence shows that the volume of M is determined by the eigenvalues. On the other hand, under suitable conditions on M, there is a Poisson relation relating the length spectrum of M to the spectrum of the Laplace operator Δ, which says roughly the singularities of the Fourier transform of the counting function of the eigenvalues λ_i are supported on the length spectrum, and is a generalization of the Poisson relation in harmonic analysis on $\mathbb{R}/\mathbb{Z}$. (See [Ch] [DG] for details).

When M is a noncompact Riemannian manifold, the spectrum of M is not discrete in general, and its structure is difficult to understand. For example, it is not easy to decide whether there are continuous spectrum and other types of spectrum, and to describe the continuous spectrum and their eigenfunctions. But for locally symmetric spaces $\Gamma\backslash X$, the spectrum of Δ consists of a discrete part and a continuous part, and the generalized eigenfunctions of the continuous spectrum can be described fairly explicitly. In this sense, locally symmetric spaces are special, important manifolds in spectral geometry.

To relate the spectrum of Δ to the geometry of noncompact M, closed geodesics are not sufficient in general. This is reasonable since the closed geodesics are not adequate to describe the geometry at infinity. We need to study geodesics that go to infinity.

When $\Gamma\backslash X = \Gamma\backslash\mathbf{H}$ is a noncompact Riemann surface, there are several typical types of non-closed geodesics:

1. Geodesics running from one cusp end to another cusp end, i.e., go to infinity through cusp neighborhoods in both directions.

2. Geodesics that go to infinity through a cusp neighborhood in one direction only.

3. Unbounded geodesics that do not go to infinity in either direction.

Geodesics that they go to infinity in only one direction exist in abundance. For example, take a geodesics $\gamma(t)$ in $\mathbf{H}$ such that as $t \to +\infty$, $\gamma(t)$ goes to $i\infty$; on the other hand, when $t \to -\infty$, $\gamma(t)$ converges to a non-rational real number. Then the image of γ in $\Gamma\backslash\mathbf{H}$ is such a geodesic. Similarly, if $\gamma(t)$ is taken to be a geodesic in $\mathbf{H}$ such that in both directions, $\gamma(t)$ converges to non-rational real numbers, then the projection of γ in $\Gamma\backslash\mathbf{H}$ belongs to the third type. On the other hand, there are only countably infinitely many geodesics of type (1).

In general, we need to restrict the types of geodesics that go to infinity. In $\Gamma\backslash\mathbf{H}$, the following two conditions on a geodesic $\gamma(t)$ are equivalent:

1. As $t \to +\infty$, $\gamma(t)$ goes to the infinity of $\Gamma\backslash\mathbf{H}$, i.e., for any compact subset $\Omega_0 \subset \Gamma\backslash\mathbf{H}$, there exists t_0 such that for $t \geq t_0$, $\gamma(t) \notin \Omega_0$.

2. Suppose that $\gamma(t)$ is of the unit speed. For $t_1, t_2 \gg 0$,
$$d_{\Gamma\backslash X}(\gamma(t_1), \gamma(t_2)) = |t_1 - t_2|.$$

Geodesics that satisfy the second condition are called eventually distance minimizing (EDM) geodesics. It can be seen that in general, these two conditions are not equivalent, and EDM geodesics are the correct ones related to the geometry at infinity.

Consider the example of $\Gamma\backslash X = \Gamma_1\backslash\mathbf{H} \times \Gamma_2\backslash\mathbf{H}$, where both factors are noncompact. Let γ_1 be a unit speed EDM geodesic in $\Gamma\backslash\mathbf{H}$ and γ_2 a unit speed closed geodesic in $\Gamma_2\backslash\mathbf{H}$. Let $\alpha, \beta > 0$ be constants such that $\alpha^2 + \beta^2 = 1$. Then $\gamma(t) = (\gamma_1(\alpha t), \gamma_2(\beta t))$ is a unit speed geodesic in $\Gamma\backslash X$ which satisfies condition (1) above but not condition (2).

The EDM geodesics of $\Gamma\backslash X$ were classified in [JM], and the boundary of various compactifications can be identified with suitable equivalence classes of EDM geodesics. Here we describe the geodesic compactification.

It is well-known that a simply connected, non-positively curved manifold M admits a geodesic compactification $M \cup M(\infty)$, where $M(\infty)$ is the set of equivalence classes of (unit speed) geodesics in M defined as follows: Two geodesics γ_1, γ_2 in M are defined to be equivalent if
$$\lim\sup_{t\to+\infty} d(\gamma_1(t), \gamma_2(t)) < +\infty.$$
For any basepoint $x_0 \in M$, let $T_{x_0}M$ be the tangent space at x_0. Then $M(\infty)$ can be identified with the unit sphere in $T_{x_0}M$ since there is a unique geodesic from x_0 belonging to each equivalence class, and hence $M(\infty)$ is called the sphere at infinity. The topology of $M \cup M(\infty)$ is defined as follows: a unbounded sequence $y_j \in M$ converges to an equivalence class $[\gamma]$ if and only if the geodesic from the basepoint x_0 to y_j converges to a geodesic in the equivalence class $[\gamma]$.

This method does not apply directly to the non-simply connected manifold $\Gamma\backslash X$. In [JM], it was modified to define the geodesic compactification $\Gamma\backslash X \cup \Gamma\backslash X(\infty)$. Specifically, $\Gamma\backslash X(\infty)$ is the set of equivalence classes of EDM geodesics in $\Gamma\backslash X$, and the topology on $\Gamma\backslash X \cup \Gamma\backslash X(\infty)$ is defined by choosing a base compact subset Ω_0 rather than a basepoint, since given any basepoint, there could exist points which are not connected to the basepoint by EDM geodesics. In this case, $\Gamma\backslash X(\infty)$ is not a sphere, rather is a finite simplicial complex which is the quotient by Γ of the rational Tits building $\Delta_{\mathbb{Q}}(\mathbf{G})$ of $\mathbf{G}$. Recall that the Tits building $\Delta_Q(\mathbf{G})$ is an infinite simplicial complex with one simplex $\sigma_{\mathbf{P}}$ for each rational parabolic subgroup $\mathbf{P}$ such that this assignment satisfies the following compatibility conditions:

1. For any pair of rational parabolic subgroups $\mathbf{P}_1, \mathbf{P}_2$, if $\mathbf{P}_1 \subset \mathbf{P}_2$, then $\sigma_{\mathbf{P}_2}$ is a simplicial face of $\sigma_{\mathbf{P}_1}$.

2. When $\mathbf{P}$ is a maximal rational parabolic subgroup, $\sigma_{\mathbf{P}}$ is a vertex, i.e., a simplex of dimension 0.

The arithmetic subgroup Γ acts on the set of rational parabolic subgroups by conjugation and hence acts on the Tits building $\Delta_{\mathbb{Q}}$ simplicially. By the reduction theory in the previous section, there are only finitely many Γ-conjugacy classes of rational parabolic subgroups, and hence the quotient $\Gamma\backslash\Delta_{\mathbb{Q}}(\mathbf{G})$ is a finite simplicial complex.

As mentioned earlier, to describe the continuous spectrum of noncompact $\Gamma\backslash X$, we need to use geodesics going to infinity. It turns out that the relevant geodesics are those which are EDM in both directions, i.e., as $t \to \pm\infty$. Hence they go from one part of infinity to another part of infinity of $\Gamma\backslash X$. They are called *scattering geodesics*. The length of scattering geodesics are clearly infinite. But we can define a finite normalization, the so-called sojourn time, which basically measures the time they spend around the compact core of $\Gamma\backslash X$, ie., Ω_0 in the decomposition given in the precise reduction theory (Proposition 6.19). When the $\mathbb{Q}$-rank of $\mathbf{G}$ is equal to 1, they are related to the the singularities of the Fourier transform of the scattering matrices of $\Gamma\backslash X$, where the generalized eigenfunctions of $\Gamma\backslash X$ are given by Eisenstein series, and the constant term of the Eisenstein series is described by the scattering matrices. See [JZ] for details.

8 Compactifications of locally symmetric spaces

In this section, we discuss several compactifications of $\Gamma\backslash X$ which arise from questions in topology of and analysis on $\Gamma\backslash X$.

A short answer to the question why we compactify $\Gamma\backslash X$ is that working with compact manifolds allows us to simplify things and to clarify structures of noncompact $\Gamma\backslash X$. We first discuss specific motivations of several compactifications and then describe their constructions briefly.

Borel-Serre compactification $\overline{\Gamma\backslash X}^{BS}$.

The first is the Borel-Serre compactification $\overline{\Gamma\backslash X}^{BS}$ of $\Gamma\backslash X$. Since X is simply connected and nonpositively curved, X is diffeomorphic to the tangent space $T_{x_0}M$ (In fact, this fact also follows directly from the Cartan decomposition). If Γ is torsion free, then

$$H^*(\Gamma, \mathbb{Z}) = H^*(\Gamma\backslash X, \mathbb{Z}).$$

This follows from a slightly stronger result that $\Gamma\backslash X$ is a $K(\Gamma, 1)$-space, or a classifying space of Γ. Recall that a $K(\Gamma, 1)$-space is a space M such that $\pi_1(M) = \Gamma$, and $\pi_i(M) = \{1\}$ for $i \geq 2$ (or the universal cover of M is contractible). If $\Gamma\backslash X$ is compact, then it is homotopic to a finite CW-complex, and hence $\Gamma\backslash X$ gives a finite classifying space. The existence of a finite classifying space of Γ implies many group theoretic finiteness properties of Γ, for example, finite generation, finite presentation etc.

When $\Gamma\backslash X$ is not compact, it is not a finite classifying space of Γ. On the other hand, if there is a compactification of $\Gamma\backslash X$ which is homotopic to $\Gamma\backslash X$ and has the homotopic type of a finite CW-complex, then the compactification gives a finite classifying space for Γ. If the compactification is a manifold with corners, then these conditions are satisfied. The Borel-Serre compactification $\overline{\Gamma\backslash X}^{BS}$ in [BS] (see also [BJ]) is a real analytic manifold with corners. When $\Gamma\backslash X = \Gamma\backslash\mathbf{H}$, a circle is added to each cusp neighborhood so that $\overline{\Gamma\backslash\mathbf{H}}^{BS}$ is a manifold with boundary, which is the union of circles.

Since $H^*(\Gamma\backslash X, \mathbb{C}) = H^*(\overline{\Gamma\backslash X}^{BS}, \mathbb{C})$, the compactification $\overline{\Gamma\backslash X}^{BS}$ allows us to decompose $H^*(\Gamma\backslash X, \mathbb{C})$ into the interior cohomology and boundary cohomology. This decomposition is related to the spectral decomposition of $\Gamma\backslash X$. In fact, automorphic forms play an important role in the study of cohomology groups of Γ (see [BW]).

Reductive Borel-Serre compactification.

The DeRham theorem identifies the singular cohomology groups of a compact manifold with the DeRham cohomology groups defined through the complex of differential forms, and the Hodge theorem picks out a canonical harmonic representative in each cohomology class. When the manifold is noncompact (and complete), a natural generalization is the L^2-cohomology which is defined as the cohomology of L^2-differential forms, and each L^2-cohomology class contains a unique harmonic representative. In the study of the DeRham cohomology, the partition of unity associated with any finite cover plays an important role. The Borel-Serre compactification is too large and does not admit partition of unity for most finite covers such that the functions have bounded derivatives with respect to the invariant metric. The problem is that in the horospherical decomposition $X = N_{\mathbf{P}} \times A_{\mathbf{P}} \times X_{\mathbf{P}}$ for any rational parabolic subgroup $\mathbf{P}$, the norm of the differential in the $N_{\mathbf{P}}$-component goes to infinity when the $A_{\mathbf{P}}$-component goes to infinity. To overcome this difficulty, we need to blow down the $N_{\mathbf{P}}$-part of the boundary components of the Borel-Serre compactification to get the reductive Borel-Serre compactification. The reductive Borel-Serre compactification turns out to be the natural compactification for L^p-cohomology of $\Gamma\backslash X$ as well (see [Zu2] [Zu3]).

Baily-Borel compactification of Hermitian locally symmetric spaces.

Assume that $X = G/K$ is a Hermitian symmetric space, i.e., a symmetric space with a G-invariant complex structure. If X is of noncompact type, then X is biholomorphic to a bounded symmetric domain. Then $\Gamma\backslash X$ is a complex space, and one question concerns the transcendental degree of the field of meromorphic functions, which was first studied by Siegel. Since quotients of holomorphic modular forms on X with respect to Γ of the same weight are meromorphic functions on $\Gamma\backslash X$, this question is related to the growth of the dimension of the space of holomorphic modular forms.

If $\Gamma\backslash X$ is compact, then it is known by the Kodaira's embedding theorem that $\Gamma\backslash X$ is a projective variety, and hence the transcendental degree is equal to the complex dimension of $\Gamma\backslash X$.

Assume that $\Gamma\backslash X$ is noncompact. If $\Gamma\backslash X$ admits a compactification $\overline{\Gamma\backslash X}$ which is a normal projective variety and the codimension of the boundary $\overline{\Gamma\backslash X} - \Gamma\backslash X$ is of complex codimension at least 2, then the Riemann extension theorem says that every meromorphic function on $\Gamma\backslash X$ extends to a meromorphic function on the projective variety $\overline{\Gamma\backslash X}$, and hence the transcendental degree is also equal to the complex dimension of $\Gamma\backslash X$.

Such a compactification is given by the Baily-Borel compactification in [BB]. In general, the compactification $\overline{\Gamma\backslash X}^{BB}$ is singular. When Γ is torsion free, the singular locus is equal to $\overline{\Gamma\backslash X}^{BB} - \Gamma\backslash X$, which is a union of lower dimensional Hermitian locally symmetric spaces.

The usual (singular) cohomology of singular spaces does not satisfy the Poincare duality in general. For complex varieties, there is a canonical intersection cohomology. It turns out that the intersection cohomology of the Baily-Borel compactification $\overline{\Gamma\backslash X}^{BB}$ is isomorphic to the L^2-cohomology of $\Gamma\backslash X$, which is called the Zucker conjecture and proved by Saper-Stern and Loojenga (see [Sap1]).

As mentioned in the previous section, the great Picard theorem on essential singularities can be stated in terms of hyperbolic embeddings. It was proved by Borel [Bo3] that the embedding of $\Gamma\backslash X$ into $\overline{\Gamma\backslash X}^{BB}$ is a hyperbolic embedding. A simpler proof was given in [Ji2]. It follows that any holomorphic map from the punctured disc D^* to $\Gamma\backslash X$ extends to a holomorphic map from D to $\overline{\Gamma\backslash X}^{BB}$. It also follows that for any compactification $\overline{\Gamma\backslash X}$ of $\Gamma\backslash X$ which is a complex space and whose boundary added at infinity is a divisor with normal crossing dominates $\overline{\Gamma\backslash X}^{BB}$, i.e., the identity map on $\Gamma\backslash X$ extends to a continuous (and hence holomorphic) map. This means in some sense that $\overline{\Gamma\backslash X}^{BB}$ is a minimal complex compactification. The Baily-Borel compactification is a Satake compactification as a topological space. In fact, there are finitely many Satake compactifications, which are topological compactifications and are partially ordered, and the Baily-Borel compactification is one of the minimal elements in this partially ordered set.

Satake compactifications.

Satake initiated the modern study of compactifications of symmetric and locally symmetric spaces in [Sat1] [Sat2]. He first constructed compactifications $\overline{X}^S$ of symmetric spaces. There are finitely many non-isomorphic Satake compactifications, which form a partially ordered set. Then he decomposed the boundary of $\overline{X}^S$ into boundary components parametrized by certain collections of real parabolic subgroups. To construct compactifications of $\Gamma\backslash X$, he defined rational boundary components and a new topology, called the Satake topology, on the union of X and the rational boundary components and showed that Γ acts continuously on this partial compactification of X with a compact quotient. Both the definition of the rational boundary components and the Satake topology depend on a choice of fundamental set. As mentioned earlier, the Baily-Borel compactification is one of the minimal Satake compactifications. In general, Satake compactifications are only topological spaces.

Toroidal compactifications.

By the Hironaka resolution theorem, the singularities of the Baily-Borel compactification $\overline{\Gamma\backslash X}^{BB}$ can be resolved. But it is desirable to get explicit resolutions. The toroidal compactifications were constructed in [AMRT] to resolve the singularities of $\overline{\Gamma\backslash X}^{BB}$. In general, when the $\mathbb{R}$-rank is greater than or equal to 2, there are infinitely many of toroidal compactifications $\overline{\Gamma\backslash X}_{\Sigma}^{tor}$, parametrized by certain polyhedral cone decompositions Σ. Each such compactification has at worst toric singularities, and infinitely many of them are smooth projective varieties. For example, there are infinitely many toroidal compactifications of the Hilbert modular surfaces, though there is only one toroidal compactification for the Picard modular surface. Torus embeddings, or toric varieties, play an important role in constructing these compactifications. In fact, suitable covering spaces of $\Gamma\backslash X$ are contained in torus bundles, and these polyhedral cones were needed to define torus embeddings of the fibers.

Relations between different compactifications.

It was proved by Zucker [Zu1] that the Borel-Serre compactification and the reductive Borel-Serre compactification $\overline{\Gamma\backslash X}^{BS}$ dominate all the Satake compactifications, in particular, the Baily-Borel compactification for Hermitian locally symmetric spaces. On the other hand, by construction, the toroidal compactifications $\overline{\Gamma\backslash X}_{\Sigma}^{tor}$ dominate the Baily-Borel compactification. Therefore, the Baily-Borel compactification $\overline{\Gamma\backslash X}^{BB}$ is a common quotient of $\overline{\Gamma\backslash X}^{BS}$ and $\overline{\Gamma\backslash X}_{\Sigma}^{tor}$. Since the constructions of $\overline{\Gamma\backslash X}^{BS}$ and $\overline{\Gamma\backslash X}_{\Sigma}^{tor}$ are completely different, Harris and Zucker conjectured in [HZ] that $\overline{\Gamma\backslash X}^{BB}$ is the greatest common quotient of $\overline{\Gamma\backslash X}^{BS}$ and $\overline{\Gamma\backslash X}_{\Sigma}^{tor}$. This turns out to be false. In fact, the greatest common quotient is sometimes strictly greater than $\overline{\Gamma\backslash X}^{BB}$. For example, when $\Gamma\backslash X$ is a Picard modular surface, there is a unique toroidal compactification $\overline{\Gamma\backslash X}^{tor}$, which strictly dominates $\overline{\Gamma\backslash X}^{BB}$ but is strictly dominated by $\overline{\Gamma\backslash X}^{BS}$. Hence, the GCQ is not equal to $\overline{\Gamma\backslash X}^{BB}$ in this case. The GCQ is described explicitly in [Ji3] and a criterion is given which allows one to decide if the GCQ is equal to $\overline{\Gamma\backslash X}^{BB}$. For example, the GCQ is equal to $\overline{\Gamma\backslash X}^{BB}$ for Hilbert modular surfaces.

Construction of compactifications.

After recalling these different compactifications of $\Gamma\backslash X$, we describe a uniform approach to constructions of compactifications.

Let $\mathbf{P}_1, \cdots, \mathbf{P}_n$ be a set of representatives of Γ-conjugacy classes of rational parabolic subgroups. Then the precise reduction theory in the previous section says that

$$\Gamma\backslash X = \Omega_0 \cup \coprod_{i=1}^{n} \pi(U_i \times A_{\mathbf{P}_i,T_i} \times V_i).$$

Since U_i, V_i can be taken to be compact subsets, the noncompactness of $\Gamma\backslash X$ arises from the cones $A_{\mathbf{P}_i,T_i}$ in the split components $A_{\mathbf{P}_i}$. It suggests that it suffices to compactify $A_{\mathbf{P}_i,T_i}$. In general, there are problems with this approach. First, the above decomposition is not unique but depends on the choices of U_i, V_i and T_i. Second, when $A_{\mathbf{P}_i,T_i}$ is compactified to $\overline{A_{\mathbf{P}_i,T_i}}$, the product structure $U_i \times A_{\mathbf{P}_i,T_i} \times V_i$ does not extend to $U_i \times \overline{A_{\mathbf{P}_i,T_i}} \times V_i$, i.e., the map from $U_i \times \overline{A_{\mathbf{P}_i,T_i}} \times V_i$ to the compactification is not injective, and some parts of the fibers $U_i \times V_i$ need to be collapsed.

Due to these and other reasons, it is better and more common to construct a Γ-equivariant partial compactification $\overline{X}$ of X and to show that Γ acts on $\overline{X}$ continuously with a compact quotient.

Suggested by the method in [BS], a general modified approach was proposed in [BJ]:

1. Choose a Γ-invariant collection of rational parabolic subgroups.

2. For each rational parabolic subgroup $\mathbf{P}$ in the collection, choose a boundary component $e(\mathbf{P})$.

3. Attach all these boundary components $e(\mathbf{P})$ to X to get a partial compactification $X \cup \coprod_{\mathbf{P}} e(\mathbf{P})$.

4. Show that $X \cup \coprod e(\mathbf{P})$ is a Hausdorff space and Γ acts continuously with compact quotient.

In step (4), the reduction theory developed in the previous section plays a crucial role. By varying the choices of the collection of rational parabolic subgroups and their boundary components, all the compactifications mentioned earlier can be constructed using this approach.

We illustrate these steps using the example of $\overline{\Gamma\backslash X}^{BS}$. In this case, we take the whole collection of rational parabolic subgroups. For each $\mathbf{P}$, its boundary component is given by

$$e(\mathbf{P}) = N_{\mathbf{P}} \times X_{\mathbf{P}}.$$

The boundary component $e(\mathbf{P})$ is attached to the infinity of X through the horospherical decomposition $X = N_{\mathbf{P}} \times A_{\mathbf{P}} \times X_{\mathbf{P}}$ as the $A_{\mathbf{P}}$-component goes to infinity through the positive chamber associated with $\mathbf{P}$.

The Hausdorff property and the compactness of the quotient $\Gamma\backslash\overline{X}^{BS}$ follows from the reduction theory and the fact that the closure of any Siegel set in $\overline{X}^{BS}$ is a compact subset in $\overline{X}^{BS}$.

To construct $\overline{\Gamma\backslash X}^{BBS}$, we still choose the whole collection of rational parabolic subgroups, and define the boundary component of $\mathbf{P}$ to be

$$e(\mathbf{P}) = X_{\mathbf{P}},$$

i.e., the unipotent factor $N_{\mathbf{P}}$ in the boundary component of the Borel-Serre compactification is collapsed, which is necessary to obtain partitions of unity as explained above.

For the Baily-Borel compactification $\overline{\Gamma\backslash X}^{BB}$, we choose the collection of maximal parabolic subgroups of $\mathbf{G}$. Recall that in this case, X is a Hermitian symmetric space, and we want the compactification of $\Gamma\backslash X$ to be a complex space. Therefore, the boundary component $e(\mathbf{P})$ needs to be a complex space, or even to be a Hermitian symmetric space of smaller dimension. In general, for a rational parabolic subgroup $\mathbf{P}$, the boundary symmetric space $X_{\mathbf{P}}$ is not Hermitian. But $X_{\mathbf{P}}$ admits an isometric decomposition

$$X_{\mathbf{P}} = X_{\mathbf{P},h} \times X_{\mathbf{P},l},$$

where $X_{\mathbf{P},h}$ is Hermitian, and $X_{\mathbf{P},l}$ is a linear symmetric space. Take the boundary component to be $e(\mathbf{P}) = X_{\mathbf{P},h}$, and

$$\overline{X}^{BB} = X \cup \coprod_{\mathbf{P}} X_{\mathbf{P},h}$$

with suitable topology, where $\mathbf{P}$ runs over maximal rational parabolic subgroups.

The above general procedure can also be used to compactify $\Gamma\backslash G$. In the compactifications of $\Gamma\backslash X$, the boundary components $e(\mathbf{P})$ are often attached at the infinity through the horopsherical decomposition of X with respect to $\mathbf{P}$. This decomposition can be replaced by the following decomposition of G:

$$G = N_{\mathbf{P}} A_{\mathbf{P}} M_{\mathbf{P}} K \cong N_{\mathbf{P}} \times A_{\mathbf{P}} \times (M_{\mathbf{P}} K).$$

Following similar steps, we can construct the Borel-Serre compactification $\overline{\Gamma\backslash G}^{BS}$, and the reductive Borel-Serre compactification $\overline{\Gamma\backslash G}^{RBS}$. The right K-action on $\Gamma\backslash G$ extends to the compactifications, and they are related to the corresponding compactifications of $\Gamma\backslash X$ by

$$\overline{\Gamma\backslash G}^{BS}/K = \overline{\Gamma\backslash X}^{BS}, \quad \overline{\Gamma\backslash G}^{RBS}/K = \overline{\Gamma\backslash X}^{RBS}.$$

Since the period domains in the theory of variation of Hodge structures are of the form $\Gamma\backslash G/H$, where H is a non-maximal compact subgroup and compactifications of $\Gamma\backslash G/H$ were sought after, the compactifications of $\Gamma\backslash G$ can be used to define compactifications of the period domains $\Gamma\backslash G/H$ (see [Gr] [GS]).

The above constructions are intrinsic in the sense that we define the ideal boundary points and the topologies of the compactifications in terms of the internal structures of $\Gamma\backslash X$. There is another approach. For any noncompact G-space Y, find a compact G-space Z, and a G-equivariant embedding $i : Y \to Z$. Then the closure of $i(Y)$ in Z is a G-compactification, i.e., a compactification with a continuous G-action. The Satake compactifications, and the Furstenberg compactifications of symmetric spaces of X were obtained this way. This approach is direct and the ideal points can be interpreted in terms of points in

Z. On the other hand, it often takes some work to understand the topology intrinsically.

We briefly discuss a compactification of $\Gamma\backslash G$ obtained this way in [BJ]. Let $\mathcal{S}(G)$ be the space of closed subgroups of G. It is a compact space with the following topology: For any $H \in \mathcal{S}(G)$, a compact subset $C \subset G$, a small neighborhood U of e, define

$$V(H, C, U) = \{H' \in \mathcal{S}(G) \mid H' \cap C \subset U(H \cap C), H \cap C \subset U(H' \cap C)\}.$$

A neighborhood of H is any set containing some $V(H, C, U)$.

There is a natural map

$$i_\Gamma : \Gamma\backslash G \to \mathcal{S}(G), \quad \Gamma g \mapsto g^{-1}\Gamma g.$$

When Γ is equal to its own normalizer $\mathcal{N}(\Gamma)$, the map i_Γ is injective. But this does not guarantee that i_Γ is an embedding, since we need to show the compatibility of the two topologies. It turns out that when Γ is an arithmetic subgroup and $\mathcal{N}(\Gamma) = \Gamma$, then i_Γ is an embedding, and the closure of $i_\Gamma(\Gamma\backslash G)$ in $\mathcal{S}(G)$ is called the subgroup compactification and denoted by $\overline{\Gamma\backslash G}^{sb}$. The limit subgroups are conjugates of the stabilizers of the constant terms in the theory of automorphic forms. In proving these results, the reduction theory for Γ plays an important role.

The condition $\mathcal{N}(\Gamma) = \Gamma$ is satisfied when Γ is a maximal discrete subgroup, i.e., it is not contained properly in any other discrete subgroup. For example, $SL(n, \mathbb{Z})$ is a maximal discrete subgroup of $SL(n, \mathbb{R})$.

The subgroup compactification $\overline{\Gamma\backslash G}^{sb}$ is closely related to $\overline{\Gamma\backslash G}^{RBS}$. In fact, there is a G-equivariant continuous map $\overline{\Gamma\backslash G}^{RBS} \to \overline{\Gamma\backslash G}^{sb}$, which is an isomorphism in some cases, for example when $\Gamma = SL(n, \mathbb{Z})$.

To modify this approach to compactify $\Gamma\backslash X$, we define $\mathcal{S}(G)/K$ to be the space of K-orbits in $\mathcal{S}(G)$. It is a compact space. When $\mathcal{N}(\Gamma) = \Gamma$, we have an embedding

$$\Gamma\backslash X \to \mathcal{S}(G)/K, \quad \Gamma g K \mapsto K \cdot (g^{-1}\Gamma g).$$

The closure of $\Gamma\backslash X$ gives a compactification $\overline{\Gamma\backslash X}^{sb}$. It can be shown that $\overline{\Gamma\backslash X}^{RBS}$ dominates $\overline{\Gamma\backslash X}^{sb}$, and the two compactifications are isomorphic to each other under certain conditions.

9 Spectral theory of locally symmetric spaces

In this section, we briefly mention applications of the reduction theory to the spectral theory of $\Gamma\backslash X$.

Let $\mathbf{G}$ be a semisimple linear algebraic group defined over $\mathbb{Q}$, $\Gamma \subset \mathbf{G}(\mathbb{Q})$ an arithmetic subgroup. Assume that $\Gamma\backslash X$ is noncompact as above.

As mentioned earlier, it follows from the reduction theory that $\Gamma\backslash X$ has finite volume. An immediate corollary is that the constant functions are square-integrable eigenfunctions in $L^2(\Gamma\backslash X)$ of the Laplace operator. But the continuous spectrum of $\Gamma\backslash X$ is nonempty. For example, when $\Gamma\backslash X = \Gamma\backslash \mathbf{H}$, it has continuous spectrum $[\frac{1}{4}, +\infty)$ with multiplicity equal to the number of ends.

The generalized eigenfunctions of the continuous spectrum are given by Eisenstein series. There are two steps in the construction of Eisenstein series:

1. Absolute convergence of Eisenstein series when the parameter is "sufficiently positive".
2. Meromorphic continuation of the Eisenstein series to the whole complex plane.

In both these steps, the reduction theory plays an important role.

Consider the example of $\Gamma\backslash X = \Gamma\backslash \mathbf{H}$. There is one Eisenstein series for each cusp of $\Gamma\backslash \mathbf{H}$. For simplicity, we only consider the one associated with the cusp $i\infty$. Let

$$\Gamma_\infty = \{\gamma \in \Gamma \mid \gamma = \pm \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}\} = \{\gamma \in \Gamma \mid \gamma(i\infty) = i\infty\}.$$

Then the Eisenstein series $E_\infty(z, s)$ of $i\infty$ is

$$E_\infty(z, s) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} (\mathrm{Im}\gamma z)^s, \quad \mathrm{Re} s > 1,$$

where the series converges absolutely under the restriction $\mathrm{Re}s > 1$. It can be shown that this series can be meromorphically continued to $s \in \mathbb{C}$, and is holomorphic at $\mathrm{Re}s = \frac{1}{2}$, and $E_\infty(z, \frac{1}{2} + ir)$, $r \in \mathbb{R}$, are generalized eigenfunctions of the continuous spectrum

$$\Delta E_\infty(z, \frac{1}{2} + ir) = (\frac{1}{4} + r^2) E_\infty(z, \frac{1}{2} + ir).$$

The reduction theory also plays an important role in understanding the behaviors at infinity of automorphic forms, for example, through the notion of constant terms (see [Bo5]). The precise reduction theory was motivated by questions in the Arthur-Selberg trace formula.

References

[AMRT] A.Ash, D.Mumford, M.Rapoport, Y.Tai, *Smooth compactifications of locally symmetric varieties*, Math. Sci. Press, Brookline, 1975.

[AS] M.Atiyah, W.Schmid, *A geometric construction of the discrete series for semisimple Lie groups*, Invent. Math. 42 (1977) 1-62.

[BGS] W.Ballmann, M.Gromov, V.Schroeder, *Manifolds of nonpositive curvature*, Prog. in Math., vol. **61**, Birkhäuser, Boston, 1985.

[BPV] W.Barth, C.Peters, A. Van de Ven, *Compact complex surfaces*, Springer-Verlag, 1984.

[BCG] G.Besson, G.Courtois, S.Gallot, *Entropies et rigiditis des espaces localement symitriques de courbure strictement nigative*, Geom. Funct. Anal. 5 (1995) 731–799.

[Bo1] A.Borel, *Semisimple groups and Riemannian symmetric spaces*, Texts & Readings in Math. vol. **16**, Hindustan Book Agency, New Delhi, 1998.

[Bo2] A.Borel, *Introduction aux groupes arithmétiques*, Hermann, Paris, 1969.

[Bo3] A.Borel, *Some metric properties of arithmetic quotients of symmetric spaces and an extension theorem*, J. of Diff. Geom. **6** (1972), 543-560.

[Bo4] A.Borel, *Linear algebraic groups*, Proc. of Symp. in Pure Math., Vol. **9** (1969), pp. 3-19.

[Bo5] A.Borel, *Automorphic forms on* $SL_2(\mathbf{R})$, Cambridge Tracts in Mathematics **130**, Cambridge University Press, 1997.

[BJ] A.Borel, L.Ji, *Compactifications of locally symmetric spaces*, preprint.

[BW] A.Borel, N.Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, Second edition, American Mathematical Society, Providence, RI, 2000

[Ch] I.Chavel, *Eigenvalues in Riemannian geometry*, Academic Press, Inc., Orlando, FL, 1984.

[CR] B.Conrad, K.Rubin, *Arithmetic algebraic geometry*, IAS/Park City Math Series, vol. 9, AMS, 2001.

[Cox] D.Cox, *Primes of the form* $x^2 + ny^2$, Wiley-Interscience, 1989.

[DM] P.Deligne, G.Mostow, *Monodromy of hypergeometric functions and non-lattice integral monodromy*, I.H.E.S. 63 (1986) 5–90.

[Do] I.Dolgachev, *Mirror symmetry for lattice polarized* $K3$ *surfaces*, J. Math. Sci. 81 (1996) 2599–2630.

[DG] J.Duistermaat, V.Guillemin, *The spectrum of positive elliptic operators and periodic bicharacteristics*, Invent. Math. 29 (1975) 39–79.

[EGM] J.Elstrodt, F.Grunewald, J.Mennicke, *Groups acting on hyperbolic spaces*, Springer-Verlag, 1998.

[FRR] S.Ferry, A.Ranicki, J.Rosenberg, *Novikov conjectures, Index theorems and Rigidity*, vol. 1, Cambridge University Press, 1995.

[Fr] E.Freitag, *Hilbert modular forms*, Springer-Verlag, 1990.

[Ga] P.Garrett, *Holomorphic Hilbert modular forms*, Wadsworth & Brooks/Cole, 1990.

[Ge] G. van der Geer, *Hilbert modular surfaces*, Springer-Verlag, 1988.

[GGP] I.Gelfand, M.Graev, I.Pisteskii-Shapiro, *Representation theory and automorphic functions*, Saunders Company, 1968.

[Go] R.Godement, *Domaines fondamentaux des groupes arithmétiques*, Bourbaki seminar, Expose 257, 1963.

[Gol] J.Goldman, *The queen of mathematics*, A.K.Peters, 1998.

[Gr] P.Griffiths, *Periods of integrals on algebraic manifolds: Summary of main results and discussion of open problems*, Bull. Amer. Math. Soc. 76 (1970) 228-296.

[GH] P.Griffiths, J.Harris, *Principles of algebraic geometry*, J.Wiley & Sons, 1978.

[GS] P.Griffiths, W.Schmid, *Recent developments in Hodge theory: a discussion of techniques and results*, in *Discrete subgroups of Lie groups and applications to moduli* (Internat. Colloq., Bombay, 1973), pp. 31–127. Oxford Univ. Press, Bombay, 1975.

[HZ] M.Harris, S.Zucker, *Boundary cohomology of Shimura varieties II. Hodge theory at the boundary*, Invent. Math. **116** (1994), 243-307.

[He] S.Helgason, *Differential Geometry, Lie groups, and symmetric spaces*, in *Pure and Applied Math.*, vol. **80**, Academic Press, 1978

[Ho1] R.Holzapfel, *Geometry and arithmetic around Euler partial differential equations*, Reidel Publishing Company, 1986.

[Ho2] R.Holzapfel, *The ball and some Hilbert problems*, Birkhauser, 1995.

[Hu] B.Hunt, *The geometry of some special arithmetic quotients*, Lect. Notes in Math., vol. 1637, Springer-Verlag, 1996.

[Ja] N.Jacobson, *Lie algebra*, Intersc. Publ., New York, 1962.

[Ji1] L.Ji, *An introduction to symmetric spaces and their compactifications*, to appear in *Lie theory*, Birkhauser, 2003.

[Ji2] L.Ji, *Metric compactifications of locally symmetric spaces*, International J. of Math. **9** (1998), 465–491.

[Ji3] L.Ji, *The greatest common quotient of Borel–Serre and the toroidal compactifications*, Geometric and Functional Analysis, **8** (1998), 978–1015.

[JM] L.Ji, R.MacPherson, *Geometry of compactifications of locally symmetric spaces*, Ann. Inst. Fourier (Grenoble) **52** (2002), no. 2, 457–559.

[JZ] L.Ji, M.Zworski, *Scattering matrices and scattering geodesics*, Ann. Scient. de Éc. Norm. Sup. **34** (2001), 441-469.

[Ka] S.Katok, *Fuchsian groups*, University of Chicago Press, 1992.

[Kn] A.Knapp, *Representation theory of semisimple groups*, Princeton University Press, 1986.

[Kob] S.Kobayashi, *Hyperbolic manifolds and holomorphic mappings*, Marcel Dekker Inc., 1970.

[LR] R.Langlands, D.Ramakrishnan, *The zeta functions of Picard modular surfaces*, les publications CRM, Montreal, 1992.

[Lo1] E.Looijenga, *Compactifications defined by arrangements I: The ball quotient case*, preprint at arXiv:math.AG/0106228.

[Lo2] E.Looijenga, *Compactifications defined by arrangements II: Locally symmetric varieties of type IV*, preprint at arXiv:math.AG/0201218.

[MR] C.Maclachlan, A.Reid, *The arithmetic of hyperbolic 3-manifolds*, GTM vol. 219, Springer-Verlag, 2003.

[Mar] G.Margulis, *Discrete subgroups of semisimple Lie groups*, Springer-Verlag, 1991.

[Mok] N.Mok, *Metric rigidity theorems on Hermitian locally symmetric manifolds*, World Scientific, 1989.

[Mos] G.Mostow, *Strong rigidity of locally symmetric spaces*, Princeton University Press, 1973.

[Mu] D.Mumford, *Tata lectures on theta I*, in *Progress in Mathematics*, vol. 28, Birkhäuser, 1983.

[OW] M.S.Osborne, G.Warner, *The Selberg Trace Formula II: Partition, Reduction, Truncation*, Pacific Jour. Math. 106 (1983), 307-496.

[PR] V.Platonov, A.Rapinchuk, *Algebraic groups and number theory*, Academic Press, 1994.

[Ra] M.Raghunathan, *Discrete subgroups of Lie groups*, Springer-Verlag, 1972.

[Sap1] L.Saper, *L_2-cohomology of algebraic varieties*, Proc. of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990), 735–746, Math. Soc. Japan, Tokyo, 1991.

[Sap2] L.Saper, *Tilings and finite energy retractions of locally symmetric spaces*, Comment. Math. Helv. **72** (1997), 167-202.

[Sat1] I.Satake, *On representations and compactifications of symmetric spaces*, Ann. of Math. **71** (1960), 77-110.

[Sat2] I.Satake, *On compactifications of the quotient spaces for arithmetically defined discontinuous groups*, Ann. of Math. **72** (1960), 555-580.

[Su] D.Sullivan, *Infinitesimal computations in topology*, IHES 47 (1977) 269-331.

[Sie] C.Siegel, *Lectures on the geometry of numbers*, Springer-Verlag, 1989.

[Yo1] M.Yoshida, *Fuchsian differential equations: with special emphasis on the Gauss-Schwarz theory*, Friedr. Vieweg & Sohn, 1987.

[Yo2] M.Yoshida, *Hypergeometric functions, my love: modular interpretations of configuration spaces*, Friedr. Vieweg & Sohn, 1997.

[Yu] N.Yu, *Arithmetic of certain Calabi-Yau varieties and mirror symmetry*, in *Arithmetic algebraic geometry* ed. by B.Conrad and K. Rubin, IAS/Park City Math Series, Vol. 9, 2001, pp. 507-569.

[Zi] R.Zimmer, *Ergodic theory and semisimple groups*, Birkhduser, 1984

[Zu1] S.Zucker, *Satake compactifications*, Comment. Math. Helv. **58** (1983), 312-343.

[Zu2] S.Zucker, *On the reductive Borel-Serre compactification: L^p-cohomology of arithmetic groups*, Amer. J. Math. **123** (2001), 951-984.

[Zu3] S.Zucker, *L_2 cohomology of warped products and arithmetic groups*, Invent. Math. **70** (1982), 169-218.