Notes on Low Dimensional Modular Varieties

Donu Arapura

May 10, 2019

Contents

Chapter 1

Elliptic curves in a nutshell

1.1 Elliptic curves: elementary approach

Curves in the projective plane $\mathbb{P}^2_{\mathbb{C}}$ of degrees one and two are easy to understand. So the first interesting case is three. For historical reasons, these are called elliptic curves. More precisely, an elliptic curve is a nonsingular cubic in $\mathbb{P}^2$. We can ask how many "degrees of freedom" do we have to choose such a curve. First of all, a homogeneous cubic polynomial in x, y, z has 10 coefficients. However, any nonzero scalar multiple of a given polynomial determines the same curve. So the count should be reduced to $10-1$. Furthermore, we only care up to linear change of variables. More formally, we want to divide out by PGL_3, leaving only $1 = 10 - 1 - 8$ parameter for an elliptic curve. Of course, this discussion was not rigorous, but it can be made so.

Theorem 1.1.1. *After a linear change of variables, an elliptic curve (over $\mathbb{C}$) can be put into Weierstrass form, given by homogenizing*

$$y^2 = 4x^3 - ax - b \tag{1.1}$$

where a, b are constants such that

$$\Delta = a^3 - 27b^2 \neq 0$$

Proof. The reduction to

$$y^2 = x^3 + Ax + B$$

can be found in [Si, chap III, §1]. From here, a further linear change of the form $(x, y) \mapsto (cx, y)$, will put into Weierstrass form. □

The significance of the shape of the right side of (1.1) will be clear shortly. Note that Δ is the discriminant of the right side $4x^3 - ax - b$. So the condition $\Delta \neq 0$ is exactly the condition for this polynomial to have distinct roots. This is equivalent to the nonsingularity of the projective curve defined by (1.1). The Weierstrass form is not unique. If $a' = c^4 a$ and $b' = c^6 b$, then

$$y^2 = 4x^3 - a'x - b'$$

gives a curve isomorphic to (1.1) under the transformation $(x, y) \mapsto (c^2 x, c^3 x)$. This is in fact the only ambiguity. We can see that the quantity

$$j = 1728\frac{a^3}{\Delta}$$

is invariant under such a transformation. (The normalization factor 1728 is there by tradition, although unimportant for our purposes.) In fact, we have

Theorem 1.1.2. *Two elliptic curves over $\mathbb{C}$ in Weierstrass form are isomorphic if and only if their j-invariants coincide.*

This makes precise what we said above.

1.2 Elliptic curves: analytic theory

We now give an analytic description. We recall that a lattice in $\mathbb{C}$ is a subgroup spanned by a real basis.

Theorem 1.2.1. *Any elliptic curve is isomorphic (as a Riemann surface) to the quotient of $\mathbb{C}$ by a lattice. Conversely, any such quotient is an elliptic curve.*

We will explain the idea of the proof of the converse statement in the last theorem, referring to [Si] for details. Let say that two tori $\mathbb{C}/L$ and $\mathbb{C}/L'$ are isomorphic if there exists a nonzero $c \in \mathbb{C}^*$ such that $cL = L'$.

Lemma 1.2.2. *Any elliptic curve is isomorphic to one of the form $E_\tau = \mathbb{C}/L_\tau, L_\tau = \mathbb{Z} + \mathbb{Z}\tau$, with $\operatorname{Im}\tau > 0$.*

This is elementary, but we give the proof, since we will need the notation anyway.

Proof. Let

$$B = \{(u, v) \in \mathbb{C}^2 \mid u, v \ \mathbb{R}\text{-linearly independent}\}$$

be the set of real bases for $\mathbb{C}$. It is easy to see that this has two connected components

$$B^+ = \{(u, v) \mid \operatorname{Im}(u/v) > 0\},\ B^- = \{(u, v) \mid \operatorname{Im}(u/v) < 0\}$$

which correspond to positively and negatively oriented bases. Clearly any lattice is given by $\mathbb{Z}u + \mathbb{Z}v$, where $(u, v) \in B$. By switching u, v, if necessary, we can assume $(u, v) \in B^+$. Then $\mathbb{Z}u + \mathbb{Z}v = v(\mathbb{Z} + \mathbb{Z}u/v)$ gives the desired isomorphism. ∎

We define the Weierstrass $\wp$-function by

$$\wp(z,\tau) = \wp(z) = \frac{1}{z^2} + \sum_{\lambda\in L,\, \lambda\neq 0} \left(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}\right) \tag{1.2}$$

It is not hard to show that the terms are dominated by $Const|z|/|\lambda|^3$, and consequently that

Proposition 1.2.3. *The series converges uniformly on compact subsets to a holomorphic function on $\mathbb{C} - L$.*

Clearly, $\wp$ has poles at L.

Theorem 1.2.4. *The Weierstrass function is periodic with respect to L in the sense that*

$$\wp(z+\lambda) = \wp(z)$$

for $\lambda \in L$

Proof. By the previous proposition, we can differentiate (1.2) term by term to obtain

$$\wp'(z) = -2\sum_{\lambda\in L}\left(\frac{1}{(z-\lambda)^3}\right)$$

So clearly $\wp'$ is doubly periodic. Therefore

$$\wp(z+\lambda) = \wp(z) + c(\lambda)$$

for appropriate constants $c(\lambda)$. In particular, setting $z = -\lambda/2$ shows that

$$\wp(\lambda/2) = \wp(-\lambda/2) + c(\lambda)$$

However, we can see directly from (1.2), that $\wp(-z) = \wp(z)$. Therefore $c(\lambda) = 0$. □

An *elliptic function* (relative to L) is a meromorphic function on $\mathbb{C}$ which is periodic with respect to L. The theorem shows that $\wp$ is elliptic. An elliptic function can be viewed a meromorphic function on $\mathbb{C}/L$. From Liouville's theorem, we obtain

Proposition 1.2.5. *An entire elliptic function is constant.*

Theorem 1.2.6. *The Laurent expansion of $\wp$ at 0 is*

$$\wp(z) = \frac{1}{z^2} + \sum_{k=1}^{\infty}(2k+1)G_{2k+2}z^{2k}$$

where the coefficients, called Eisenstein series, are

$$G_{2k} = \sum_{\lambda\in L-\{0\}} \frac{1}{\lambda^{2k}}$$

Proof. This results from substituting

$$(z-\lambda)^{-2} - \lambda^{-2} = \lambda^{-2}[(1-z/\lambda)^{-2} - 1]$$
$$= \sum_{1}^{\infty} \frac{(k+1)z^k}{\lambda^{k+2}}$$

into (1.2). □

Theorem 1.2.7.
$$(\wp')^2 = 4\wp^3 - g_2\wp^2 - g_3$$
where
$$g_2 = 60G_4,\ g_3 = 140G_6$$

Sketch. Let
$$f(z) = (\wp')^2 - 4\wp^3 - g_2\wp^2 - g_3$$
This is clearly elliptic, and the only possible poles are at points of L. However, using the previous theorem we can calculate enough terms of the Laurent series of f to conclude that f has no poles at 0 and $f(0) = 0$. It follows that f has no singularities at all, and is therefore constant. So it must be identically 0.

□

We can now define a map $\mathbb{C}/L \to \mathbb{P}^2$ given by

$$z \mapsto \begin{cases} [\wp(z), \wp'(z), 1] & \text{if } z \notin L \\ [0, 1, 0] & \text{otherwise} \end{cases}$$

Proposition 1.2.8. *This is an embedding.*

Proof. See [Si, pp 158-159]. □

Putting the above statements together, we see that $\mathbb{C}/L$ is a cubic in $\mathbb{P}^2$ as claimed earlier.

One consequence of this representation of an elliptic curve as a torus, is that we get a natural group law on it.

1.3 Analytic theory continued: theta functions

With an eye towards higher dimensions, we want to give a different method of realizing the elliptic curve $E = \mathbb{C}/\mathbb{Z} + \mathbb{Z}\tau$ as a projective curve. We need to construct functions $f_i : E \to \mathbb{C}$ such that $p \mapsto [f_0(p), \ldots, f_n(p)] \in \mathbb{P}^n$ is well defined and gives an embedding. If we regard f_i as functions from $\mathbb{C} \to \mathbb{C}$, these would be quasiperiodic, in the sense that

$$f_i(p+\lambda) = (\text{some factor}) f_i(p), \quad \forall \lambda \in \mathbb{Z} + \mathbb{Z}\tau$$

where the factor in front is the same for all i and nonzero. The basic example is Jacobi's θ-function. This is given by the Fourier series

$$\theta(z,\tau) = \sum_{n\in\mathbb{Z}} \exp(\pi i n^2\tau + 2\pi i n z) = \sum_{n\in\mathbb{Z}} \exp(\pi i n^2 \tau)\exp(2\pi i n z)$$

Since τ is fixed, we just view it as function of z for now. Writing $\tau = x + iy$, with $y > 0$, shows that on a compact subset of the z-plane the terms are bounded by $O(e^{-n^2 y})$. So uniform convergence on compact sets is guaranteed, and we can conclude that θ holomorphic. This is clearly periodic

$$\theta(z+1) = \theta(z) \tag{1.3}$$

In addition it satifies the functional equation

$$\begin{aligned} \theta(z+\tau) &= \sum \exp(\pi i n^2\tau + 2\pi i n(z+\tau)) \\ &= \sum \exp(\pi i (n+1)^2\tau + 2\pi i(n+1)z - 2\pi i z - \pi i \tau) \\ &= \exp(-\pi i\tau - 2\pi i z)\theta(z) \end{aligned} \tag{1.4}$$

Conversely, if $f(z)$ is a holomorphic function satisfying these equations, then (1.3) yields a Fourier exansion

$$f(z) = \sum_n a_n \exp(2\pi i n z)$$

and (1.4) produces recurrence conditions on the coefficients. This can be used to show that $f(z) = a_0\theta(z)$. We get more solutions by relaxing these conditions. Let $N > 0$ be an integer, and consider the space V_N of holomorphic functions satisfying

$$\begin{aligned} f(z+N) &= f(z) \\ f(z+N\tau) &= \exp(-\pi i N^2\tau - 2\pi i N z) f(z) \end{aligned} \tag{1.5}$$

By the first equation, any function in V_N can be expanded in a Fourier series (in powers of $\exp(2\pi i/N)$), and the second equation yields recurrences which shows that the coefficients are determined by N^2 of them. In other words:

Lemma 1.3.1. $\dim V_N = N^2$.

A proof of this lemma can be found on pp 8-10 of [MT]. The discussion there gives quite a bit more information that we recall. The conditions (1.5) can be expressed as invariance under the operators

$$S_a(f)(z) = f(z+a),$$

$$T_b(f)(z) = \exp(\pi i b^2\tau + 2\pi i b z) f(z + b\tau)$$

for $a, b \in N\mathbb{Z}$. For $a, b \in \mathbb{R}$, we have the following identities

$$S_aS_b = S_{a+b},\ T_aT_b = T_{a+b}$$

$$S_a T_b = \exp(2\pi i ab) T_b S_a$$

So they generate a nonabelian group H, called a Heisenberg group, which fits into an exact sequence

$$1 \to U(1) \to H \to \mathbb{R}^2 \to 0$$

where the last map sends $S_a T_b \mapsto (a, b)$ and $U(1) \subset \mathbb{C}^*$ is the unit circle. A key fact is:

Lemma 1.3.2. *V_N is stable under the operators $S_{1/N}$ and $T_{1/N}$, and therefore under the subgroup H'_N of H generated by these operators. This subgroup fits into a sequence*

$$1 \to \mu_N \to H'_N \to (\frac{1}{N}\mathbb{Z})^2 \to 0$$

The action of H'_N on V_N is trivial on the preimage of $(N\mathbb{Z})^2$. Therefore the action factors through a finite quotient H_N of H'_N which, as an abstract group, fits into an exact sequence

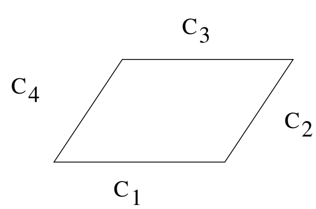
$$1 \to \mu_N \to H_N \to (\mathbb{Z}/N^2\mathbb{Z})^2 \to 0$$

Lemma 1.3.3. *Given nonzero $f \in V_N$, it has exactly N^2 zeros, counted with multiplicities, in the parallelogram with vertices $0, N, N\tau, N+\tau$ (where we translate if necessary so no zeros lie on the boundary).*

Proof. Complex analysis tells us that the number of zeros is given by the integral

$$\frac{1}{2\pi i}\int_{C_1+C_2+C_3+C_4} \frac{f'(z)dz}{f(z)}$$

over the boundary of the parallelogram.



Using $f(z+N) = f(z)$, we obtain

$$\int_{C_2+C_4} \frac{f'(z)dz}{f(z)} = 0$$

and from $f(z+N\tau) = Const.\exp(-2\pi i N z) f(z)$, we obtain

$$\int_{C_1+C_3} \frac{f'(z)dz}{f(z)} = 2\pi i N^2$$

□

A function $f \in V_N$ is quasi-periodic with respect to the lattice NL. If we transform it to $F(z) = f(Nz)$, then it become quasi-periodic with respect to L. Once we choose a basis f_i of V_N, the map $\phi : E \to \mathbb{P}^{N^2-1}$ given by $z \mapsto [F_i(z)]$ is well defined. Recall that we found a finite group H_N which acts on V_N, and therefore on $\mathbb{P}^{N^2-1}$. This group also acts on E where the image (a, b) under the homomorphism $H_N \to (\mathbb{Z}/N^2)^2$ sends $z \mapsto z + a/N^2 + b\tau/N^2$. One checks by a calculation [MT, p 13] that

Lemma 1.3.4. *ϕ is equivariant for these actions*

Theorem 1.3.5. *$\phi : E \to \mathbb{P}^{N^2-1}$ is an embedding.*

Proof. Suppose that ϕ is not one to one. Then $F(z_1) = \lambda F(z_1')$ for some $z_1 \neq z_1'$ in $\mathbb{C}/L$, some $\lambda \in \mathbb{C}^*$, and all $f \in V_N$. By translation by H_N, we can find another such pair z_2, z_2' with this property, such that z_1, z_1', z_2, z_2' are distinct. Choose $N^2 - 3$ additional points $z_3, \ldots z_{N^2-1}$ in $\mathbb{C}/L$ distinct from the previous choices. We define a map $V_N \to \mathbb{C}^{N^2-1}$ by $f \mapsto (F(z_i))$. Since $\dim V_N = N^2$, we can find a nonzero $f \in V_N$ so that

$$F(z_1) = F(z_2) = F(z_3) = \ldots F(z_{N^2-1}) = 0$$

Notice that we are forced to also have $F(z_1') = F(z_2') = 0$ which means that f has at least $N^2 + 1$ zeros which contradicts the lemma.

A similar argument shows that the derivative $d\phi$ is nowhere zero. Otherwise we would have a point z_1 such that F' has a zero at z_1 for every $f \in V_N$. Arguing as above, we would find a nonzero $f \in V_N$ and points $z_1, \ldots z_{N^2-1}$, such that F has zeros at the z_i and double zeros at z_1, z_2. This again yields a contradiction. $\square$

The embeddings produced this way are different from the previous method. The smallest case is when $N = 2$. Then we get an embedding E into $\mathbb{P}^3$. One can show that it an intersection of two quadrics. In general, we can always guarantee that the image is algebraic by:

Theorem 1.3.6 (Chow)**.** *If $X \subset \mathbb{P}^n$ is a complex submanifold, then it automatically a nonsingular projective algebraic variety.*

1.4 Elliptic curves over arbitrary fields

Finally let us redo parts of the theory assuming Hartshorne level algebraic geometry.[1] We now work over an arbitrary field k, which is not necessarily algebraically closed. An elliptic curve over k, is a smooth projective curve E over k, of genus one, with a fixed k-rational point O. We will deduce the earlier description as a consequence. First, we should recall that the fundamental invariant of a smooth projective curve is its genus g. Suppose that k is algebraically closed.

[1] And if you haven't read it, don't worry about it too much. All of this material can be understood with only basic AG, as in [Si].

If $X \subset \mathbb{P}^N$ is a smooth projective curve, let X' be its image under a general linear projection $\mathbb{P}^N \dashrightarrow \mathbb{P}^2$. Then $X' \subset \mathbb{P}^2$ has only nodes as singularities. The genus is given by

$$g = \frac{(d-1)(d-2)}{2} - \delta$$

where d is the degree of X' and δ is the number of nodes. So $g = 1$ when $X = X'$ is smooth of degree 3. Although this gives a method for calculating g, it does not make a good definition, as it not obviously independent of the choice of embedding. A more intrinsic definition is via sheaf cohomology

$$g = \dim H^1(X, \mathcal{O}_X) = \dim H^0(X, \Omega^1_X)$$

where the last equality is a special case of Serre duality. This shows that $g \geq 0$, which wasn't obvious from the last formula. When $k = \mathbb{C}$, g can also be identified with one half the first Betti number. This can be seen from the Hodge decomposition.

Among other things, the genus enters into the statement of the Riemann-Roch theorem, which we will recall. Let us suppose that k is algebraically closed for simplicity, then a divisor D is a finite formal sum $D = \sum n_i p_i$, where p_i are points of X. Define the degree

$$\deg D = \sum n_i$$

The formalism works over nonalgebraically closed fields, but now p_i are closed points of X viewed as a scheme, and $\deg D = \sum n_i [k(p_i) : k]$, where $k(p_i)$ are the residue fields. If f is a nonzero rational function, the associated principal divisor

$$\operatorname{div} f = \sum ord_p(f) p$$

where $ord_p(f)$ is the discrete valuation attached to p. If ω is a nonzero rational differential form, the canonical divisor

$$\operatorname{div} \omega = \sum ord_p(\omega)$$

In spite of the formal similarity canonical divisors are usually not principal. In fact the degrees

$$\deg(\operatorname{div} f) = 0, \ \deg(\operatorname{div} \omega) = 2g - 2$$

are usually different. However, when $g = 1$, we do have equality. In fact, more is true.

Lemma 1.4.1. *When $g = 1$, $\Omega^1_X \cong \mathcal{O}_X$ and any canonical divisor is principal.*

Proof. Since $H^0(\Omega^1_X) \neq 0$, we have a nonzero regular 1-form ω. Note that ω has no poles, and since $\deg \operatorname{div} \omega = 0$, it has no zeros either. By identifying $H^0(X, \Omega^1_X) \cong Hom(\mathcal{O}_X, \Omega^1_X)$, we can view ω as a nonzero morphism $\mathcal{O}_X \to \Omega^1_X$. The map is injective, because the kernel consists of functions f such that $f\omega = 0$. Since for every $p \in X$, $\omega(p) \neq 0$, we can express dx as multiple of ω, where x is local uniformizer. This implies that ω is surjective as well. Therefore $\Omega^1_X \cong \mathcal{O}_X$, and the second statement is an immediate consequence. $\square$

We define a sheaf $\mathcal{O}_X(D)$, sometimes denoted by $\mathcal{L}(D)$, whose global sections are

$$H^0(X, \mathcal{O}_X(D)) \cong \{f \text{ a rational function} \mid ord_{p_i} f + n_i \geq 0\}$$

This means that $\operatorname{div}(f) + D$ is effective in the sense that its coefficients are nonnegative. There are a few cases where this can be computed directly from the definition. If $D = 0$, then $H^0(\mathcal{O}_X(D)) = H^0(\mathcal{O}_X)$ consists of constant functions (because X is projective). If D is nonzero with positive coefficients, then $H^0(\mathcal{O}_X(-D))$, consists of constant functions vanishing somewhere, so that $H^0(\mathcal{O}_X(-D)) = 0$. For more general cases, we can use Riemann-Roch.

Theorem 1.4.2 (Riemann-Roch)**.**

$$h^0(\mathcal{O}_X(D)) - h^0(\mathcal{O}_X(K - D)) = \deg D + 1 - g$$

where K is any canonical divisor and $h^i = \dim H^i$.

This leads to another useful method for computing the genus.

Corollary 1.4.3. $\deg K = 2g - 2$

Proof. Apply Riemann-Roch when $D = K$. □

Let us return to the case of an elliptic curve (E, O). Take $D = nO$ (O not 0), where n is a positive integer. Then

$$h^0(\mathcal{O}_E(nO)) - h^0(\mathcal{O}_E(K - nO)) = 1$$

Since $K = 0$, we can write the second term on the left as $h^0(\mathcal{O}(-nO))$, but this is 0. Thus we can conclude that

$$h^0(\mathcal{O}_E(nO)) = n$$

It follows that $H^0(\mathcal{O}_E(O)) = H^0(\mathcal{O}_E)$ consists of just the constant functions. This also implies that there exists a nonconstant $f \in H^0(\mathcal{O}(2E))$ and a function $g \in H^0(\mathcal{O}(3O))$ not in $H^0(\mathcal{O}(2O))$. We can also conclude that the seven functions $1, f, f^2, f^3, g, g^2, gf \in H^0(\mathcal{O}(6O))$ are linearly dependent. Using these facts, it is not difficult to show that the map $p \mapsto (f(p), g(p))$ extends to embedding of E as a cubic in $\mathbb{P}^2_k$. More generally:

Theorem 1.4.4. *Suppose that D is a divisor of degree 3 or more, and $f_0, \dots f_n \in H^0(E, \mathcal{O}_E(D))$ is a basis. Then the map $\phi : E \to \mathbb{P}^n_k$ given $\phi(x) = [f_0(x), \dots, f_n(x)]$ is an embedding.*

Proof. This follows from [H, cor 3.2, p 308]. □

Recall that the class group $Cl(X)$ of a smooth projective curve is the quotient of the group of divisors by the subgroup of principal divisors. Since principal divisors have degree 0, the degree homomorphism factors through $Cl(X)$. Let $Cl^0(X) = \ker \deg Cl(X) \to \mathbb{Z}$.

Theorem 1.4.5. *Let (E, O) be an elliptic curve. The map $\alpha : E \to Cl^0(E)$ defined by $\alpha(p) = p - O$ is a bijection.*

Proof. Suppose D is divisor of degree 0. By Riemann-Roch

$$h^0(\mathcal{O}(D + O)) = 1$$

Choose a nonzero function $f \in H^0(\mathcal{O}(D + O)$. $\operatorname{div} f$ is necessary of the form p for some $p \in X$. So that $f \in H^0(\mathcal{O}(D + O - p))$ This implies that divisor class of D and $p - O$ are equal. Therefore α is surjective.

Suppose that $\alpha(p) = \alpha(q)$ and that $p \neq q$. Then $p - q$ is principle. This implies that there is a function f with simple pole at p and no other poles. Viewing f as map $f : E \to \mathbb{P}^1$, we can see that this implies that f is degree 1. So we are forced to conclude that $E \cong \mathbb{P}^1$ but this is impossible since the genera are different. So α is injective. □

Corollary 1.4.6. *E has the structure of abelian group in a natural way.*

Without the word "natural", the result would be quite useless. We can interpret this to mean, that the group operations are connected to the structure of E as an algebraic variety, in the sense that they are morphisms. We refer to [Si] or other standard texts for an explanation or why this holds.

Let us return to case when $k = \mathbb{C}$ and reinterpret the theory of theta functions in terms of divisors. Given V_N as before, $f \in V_N - \{0\}$ is *not* a function on $E = \mathbb{C}/\mathbb{Z} + \mathbb{Z}\tau$. However, we can attach an effective divisor D_f to it by taking the divisor of zeros of f in a fundamental parallelogram as in lemma 1.3.3. This lemma shows that $\deg D_f = N^2$. If $g \in V_N$ is another nonzero function, g/f is invariant and therefore a meromorphic function on E. We can see that $D_g = D_f + \operatorname{div}(g/f)$, so that D_g is linearly equivalent to D_f. This tells us that $g/f \in H^0(E, \mathcal{O}(D_f))$. So the map $g \mapsto g/f$ gives an injective homomorphism, which we can view as an inclusion

$$V_N \subseteq H^0(E, \mathcal{O}(D_f))$$

Since both sides have dimension N^2, we must have equality. In particular, theorem 1.3.5 follows from theorem 1.4.4. Finally, we note that there is even an analogue of the Heisenberg group due to Mumford. We won't get into that here, but instead refer to his paper *On the equations defining Abelian varieties I, Inventiones 1966* for details.

Chapter 2

Modular curves

2.1 The action of $SL_2(\mathbb{R})$

We let $GL_2(\mathbb{C})$ act on the Riemann sphere $\mathbb{C} \cup \{\infty\}$ by fractional linear transformations

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau \mapsto \frac{a\tau + b}{c\tau + d}$$

Note that we can identify $\mathbb{C} \cup \{\infty\} = \mathbb{P}^1_{\mathbb{C}}$. With respect to this, the above action of $GL_2(\mathbb{C})$ coincides with the usual action on the projective line by $[v] \mapsto [Av]$.

Lemma 2.1.1. *$SL_2(\mathbb{R})$ acts transitively on the upper half plane $\mathbb{H} = \{\tau \in \mathbb{C} \mid \operatorname{Im} \tau > 0\}$ by fractional linear transformations. The stabilizer of i is $SO(2)$. Therefore, we can identify $\mathbb{H} = SL_2(\mathbb{R})/SO(2)$.*

Proof. If $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R})$ and $\tau \in \mathbb{C}$, then

$$\operatorname{Im} \frac{a\tau + b}{c\tau + d} = \frac{\operatorname{Im} \tau}{|c\tau + d|^2} \tag{2.1}$$

This shows that $SL_2(\mathbb{R})$ preserves $\mathbb{H}$. We have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot i = \frac{(ca + db) + i}{c^2 + d^2}$$

It is now an easy exercise to see that given $\tau \in \mathbb{H}$, we can find a solution to

$$A \cdot i = \tau$$

with $A \in SL_2(\mathbb{R})$, and that if $\tau = i$, we must have $A \in SO(2)$.

□

We can view $\mathbb{H}$ as the upper hemisphere of the Riemann sphere $\mathbb{P}^1_{\mathbb{C}}$. The action of $SL_2(\mathbb{R})$ extends to the boundary $\partial\mathbb{H} = \mathbb{P}^1_{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. In order to better visualize the action, it useful to note that $\mathbb{H}$ has a Riemannian metric, called the hyperbolic or Poincaré metric, where the geodesics are lines or circles meeting $\partial\mathbb{H}$ at right angles. The action of $SL_2(\mathbb{R})$ preserves this metric, so it takes a geodesic to another geodesic.

2.2 The modular group $SL_2(\mathbb{Z})$

Let $L_\tau = \mathbb{Z}+\mathbb{Z}\tau$ with $\tau \in \mathbb{H}$ as before. We can see that elliptic curves $E_\tau = \mathbb{C}/L_\tau$ and $E_{\tau'}$ are isomorphic if and only if $L_\tau = L_{\tau'}$.

Lemma 2.2.1.

(a) If $(u,v)^T, (u',v')^T \in B^+$, then $\mathbb{Z}u + \mathbb{Z}v = \mathbb{Z}u' + \mathbb{Z}v'$ if and only if $(u,v)^T$ and $(u',v')^T$ lie in the same orbit of $SL_2(\mathbb{Z})$.

(b) $L_\tau = L_{\tau'}$ if and only τ, τ' lie in the same orbit under $SL_2(\mathbb{Z})$.

Proof. If $\mathbb{Z}u + \mathbb{Z}v = \mathbb{Z}u' + \mathbb{Z}v'$, there would be change of basis matrix A taking $(u,v)^T$ to $(u',v')^T$. A is necessarily integral with positive determinant, and this already ensures that $A \in SL_2(\mathbb{Z})$. The converse is easy. (b) follows from (a). □

From this lemma, we can conclude that:

Theorem 2.2.2. *The set of isomorphism classes of elliptic curves (over $\mathbb{C}$) is parameterized by $SL_2(\mathbb{Z})\backslash\mathbb{H}$.*

At the moment, $A_1 = SL_2(\mathbb{Z})\backslash\mathbb{H}$ is just a set. In order to give more structure, we need to analyze the action more carefully. First observe that $-I$ acts trivially on $\mathbb{H}$, so the action factors through $\Gamma = PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/\{\pm I\}$. Consider the closed region $F \subset \mathbb{C}$ bounded by the unit circle and the lines $\operatorname{Im} z = \pm 1/2$ depicted below.

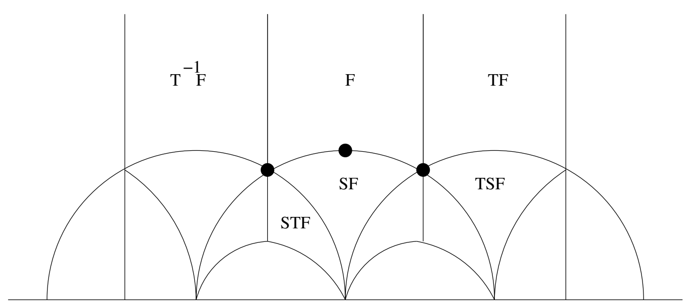


Figure 2.1: Fundamental domain

Let $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. These act by $z \mapsto -1/z$ and $z \mapsto z+1$ respectively. S is a reflection about i which interchanges the regions $|z| \geq 1$ and $|z| \leq 1$. They generate a subgroup $G \subseteq \Gamma$.

Theorem 2.2.3.

(a) The union of translates gF, $g \in G$, covers $\mathbb{H}$.

(b) An interior point of F does not lie in any other translate of F under Γ.

(c) The isotropy group of $z \in F$ is trivial unless it is one of the points $\{i, e^{\pi i/3}, e^{2\pi i/3}\}$ marked in the diagram. The isotropy group is $\langle S\rangle, \langle ST\rangle, \langle TS\rangle$ respectively.

Proof. The intuition behind this can be understood from the picture. Repeatedly applying S and $T^{\pm 1}$ to F gives a *tiling* of $\mathbb{H}$ by hyperbolic triangles. Choose $\tau \in \mathbb{H}$, we want to find $A' \in SL_2(\mathbb{Z})$ and $\tau' \in F$ such that $A' \cdot \tau' = \tau$. Using (2.1), we can see that $\{\operatorname{Im} A \cdot \tau \mid A \in SL_2(\mathbb{Z})\}$ has a maximum M. Choose an A which realizes this maximum. Choose an integer n so that $\tau' = T^n A\tau$ has real part in $[-1/2, 1/2]$. Observe that $\operatorname{Im}\tau' = M$. If $|\tau'| < 1$ then $-1/\tau'$ would have imaginary bigger than M which is impossible. It follows that $\tau' \in F$, and τ lies in its orbit. This proves (a). For the remaining parts, see Serre [Se, pp 79].

□

The set F is called a *fundamental domain* for the action of G. We can draw a number of useful conclusions.

Corollary 2.2.4. *$G = PSL_2(\mathbb{Z})$, i.e. S and T generate $PSL_2(\mathbb{Z})$.*

Proof. Let $z \in F$ be an interior point, and $h \in \Gamma$. Then $hz = gz$ for some $g \in G$. Since $z \in h^{-1}gF$, we must have $h^{-1}g = I$. □

Corollary 2.2.5. *The nontrivial elements of finite order in Γ are conjugate to S or $(ST)^{\pm 1}$.*

Proof. A nontrivial element of finite must lie in the isotropy group of some point in $\mathbb{H}$. The points in the plane with nontrivial isotropy groups must be a translate of i or $e^{2\pi i/3}$. Their isotropy groups must be conjugate to the isotropy groups of one these two points. □

Corollary 2.2.6. *The action of $PSL_2(\mathbb{Z})$ is properly discontinuous, which means that for every point $p \in \mathbb{H}$, there is a neighbourhood U such that $gU \cap U = \emptyset$ for all but finitely many g.*

We can give A_1 the quotient topology where $U \subseteq A_1$ is open if and only its pullback to $\mathbb{H}$, under the projection $\pi : \mathbb{H} \to A_1$ is open.

Proposition 2.2.7. *The topology on A_1 is Hausdorff. In fact, it is homeomorphic to $\mathbb{C}$*

Proof. The first statement follows immediately from the last corollary. Using the above results, one can see that A_1 is obtained by gluing the two bounding lines of F and folding the circular boundary in half. This is easily seen to be homeomorphic to the sphere minus the north pole. □

A_1 has a natural compactification $\bar{A}_1$ given by adding single point at infinity to make it a sphere. We will follow the convention of the automorphic form literature and call it a *cusp*. It is important to keep in mind that this clashes with the usual terminology in algebraic geometry, that a cusp is a singularity of the form $y^2 = x^3$. We will refer the last thing as cuspidal singularity in order to avoid confusion. We can construct this a quotient as follows. Let $\mathbb{H}^* = \mathbb{H} \cup \mathbb{Q} \cup \{\infty\} \subset \mathbb{P}^1$. The action of Γ on $\mathbb{P}^1$ stabilizes $\mathbb{H}^*$. On $\mathbb{H}$ it coincides with the standard action, and on $\mathbb{Q} \cup \{\infty\}$ it consists of a single orbit. Thus $\Gamma \backslash \mathbb{H}^* = \bar{A}_1$ as a set. In order to get the correct topology on the quotient, one needs a somewhat exotic topology of $\mathbb{H}^*$. On $\mathbb{H}$ it's the usual one, but on $\partial \mathbb{H}^*$ a fundamental system of punctured neighbourhoods of (a translate of) ∞ are (translates of) strips $\operatorname{Im} z > n$, $n \in \mathbb{N}$. These can be visualized as interiors of circles tangent to the boundary circle $\partial \mathbb{H}$.

2.3 Modular forms

Since A_1 has a topology, we can talk about continuous functions on it. We can see that $f : A_1 \to \mathbb{C}$ is continuous if and only if it's pullback $\pi^* f := f \circ \pi$ is continuous. Let us also declare that a function on an open subset of $\mathcal{A}_1$ is holomorphic or meromorphic if its pullback to $\mathbb{H}$ has the same property. This means that such functions correspond to Γ-invariant functions on $\mathbb{H}$. Before constructing nontrivial examples, we want to relax the condition. We say that f is automorphic, with automorphy factor $\phi_\gamma(z)$, if it satisfies the functional equation

$$f(\gamma z) = \phi_\gamma(z) f(z)$$

This is very similar to what we did with theta functions. If we have two such functions with the same factor, their ratio would be invariant. Note that for this to work, we need to impose a consistency condition

$$\phi_{\gamma\xi}(z) f(z) = f(\gamma \xi z)$$

$$= \phi_\gamma(\xi z) f(\xi z) = \phi_\gamma(\xi z) \phi_\xi(z) f(z)$$

Cancelling f, leads to a so called cocycle condition on the automorphy factor

$$\phi_{\gamma\xi}(z) = \phi_\gamma(\xi z) \phi_\xi(z)$$

As the terminology suggests, ϕ_γ does give an element of a certain cohomology group. Rather than pursuing this direction, let us look for natural automorphic forms/factors in nature. Given a meromorphic differential form $\omega = f(z)dz$ on $\mathbb{H}$, let us see how it transforms under $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$. We can see that

$$\omega \mapsto f(\gamma \cdot z) d\left(\frac{az+b}{cz+d}\right) = (cz+d)^{-2} f(\gamma \cdot z) dz$$

We say that $f(z)$ is a weakly modular form of weight 2, with respect to Γ, if $f(z)dz$ is invariant. We say that f is *weakly modular of weight* $2k$ if it

$$f(z) = (cz+d)^{-2k} f\left(\frac{az+b}{cz+d}\right) \tag{2.2}$$

This means that the tensor $f(z)dz^{\otimes k}$ is invariant. More generally, it makes sense to consider weakly modular forms of arbitrary integer weight ℓ, satisfying

$$f(z) = (cz+d)^{-\ell} f\left(\frac{az+b}{cz+d}\right)$$

However, when ℓ is odd, taking $\gamma = -I$, shows that $f = -f$, so it's zero! Natural nonzero examples do exist for other groups however, as we shall see shortly.

To drop the "weakly", we impose holomorphy conditions on $\mathbb{H}$ but also at infinity. To understand what the last part means, we first note that by using S and T, (2.2) is equivalent to

$$\begin{aligned} f(z+1) &= f(z) \\ f(-1/z) &= z^k f(z) \end{aligned} \tag{2.3}$$

The first condition means that we have a Fourier expansion

$$f(z) = \sum_{-\infty}^{\infty} a_n e^{2\pi i n z} = \sum_{-\infty}^{\infty} a_n q^n$$

where $q = e^{2\pi i z}$. Note that as $z \to i\infty$, $q \to 0$. So we want to think of q as the local parameter at infinity. Then the Fourier series becomes the Laurent series in q. f is a *modular form of weight* $2k$ if it is holomorphic in $\mathbb{H}$, (2.2) holds, and the Fourier coefficients $a_n = 0$ for $n < 0$. It is called a *cusp form* of weight $2k$ if in addition $a_0 = 0$.

Theorem 2.3.1. *The Eisenstein series*

$$G_{2k}(z) = \sum_{\mathbb{Z}^2 - 0} \frac{1}{(mz+n)^{2k}}$$

is a modular form of weight $2k$*, when* $k \geq 2$*.*

$$\Delta(z) = (60G_4(z))^3 - 27(140G_6(z))^2$$

is a cusp form of weight 12*.*

Proof. The sum can be seen to converge uniformly on compact sets, so it must converge to a holomorphic function on $\mathbb{H}$. One has

$$G_{2k}\left(\frac{az+b}{cz+d}\right) = (cz+d)^{2k} \sum \frac{1}{(ma+ndc)z + (mb+nd))^{2k}}$$

The vectors $(ma + ndc, mb + nd)$ can be seen to run over $\mathbb{Z}^2 - 0$. So the right side can be rewritten as

$$(cz + d)^{2k} G_{2k}(z)$$

as required.

We have to check holomorphy at infinity. By uniform convergence, we can evaluate the limit as $z \to \infty$ term by term. When $m \neq 0$, have $(mz+n)^{-2k} \to 0$ as $z \to \infty$. Therefore

$$\lim_{z\to\infty} G_{2k}(z) = 2\sum_{n=1}^{\infty} \frac{1}{n^{2k}} = 2\zeta(2k)$$

where ζ is the Riemann zeta function. Euler gave explicit formulas for the values

$$\zeta(4) = \frac{\pi^4}{90}$$

$$\zeta(6) = \frac{\pi^6}{945}$$

This allows us to evaluate $\lim_{z\to\infty} \Delta(z)$ and check that it's zero. □

Corollary 2.3.2.

$$j(z) = 1728 \frac{(60G_4(z))^3}{\Delta}$$

is weakly modular of weight 0*.*

Finally, let us consider Jacobi's theta function. This is a function of two variables $\theta(z, \tau)$. We already studied the behaviour in the first, now we consider the second where we set $z = 0$.

$$\theta(0, \tau) = \sum_{n\in\mathbb{Z}} \exp(\pi i n^2 \tau)$$

From this formula, we see that

$$\theta(0, \tau + 2) = \theta(0, \tau)$$

There is also a somewhat subtler functional equation.

Theorem 2.3.3. *We have*

$$\theta(0, -1/\tau) = \sqrt{-i\tau}\, \theta(0, \tau)$$

where the complex square root needs to be handled with the usual care.

Sketch. We need the Poisson summation formula [DM], which tells us that

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n)$$

where f is a rapidly decreasing smooth (aka Schwartz) function, and

$$\hat{f}(v) = \int_{-\infty}^{\infty} f(u)e^{-2\pi iuv}du$$

is its Fourier transform. The Fourier transform of the Gaussian $e^{-\pi u^2\tau}$ is $\tau^{-1/2}e^{-\pi v^2/\tau}$. Therefore the Poisson summation formula shows that

$$T(1/y) = \sqrt{y}T(y)$$

where $T(y) = \theta(0, iy)$. The theorem follows by analytic continuation. □

Corollary 2.3.4.

$$\theta(0, -1/\tau)^2 = -i\tau\theta(0, \tau)^2$$

The last equation plus the previous periodicity suggests that $\theta(0, 1/\tau)^2$ is a modular form of some kind. In fact, it is a modular form of weight one for a subgroup $\Gamma(4)$ to be defined below. See [MT, p 39].

2.4 Modular curves

With the topology of $X(1) = \bar{\mathcal{A}}_1$ constructed earlier, which is homeomorphic to $\mathbb{P}^1$, we can construct a sheaf of functions $\mathcal{O}_{X(1)}$ as follows. Let $\Gamma(1) = SL_2(\mathbb{Z})$. Given a $\Gamma(1)$-invariant open set $\tilde{U} \subset \mathbb{H}$, let us say that a holomorphic function f on it is modular of weight $2k$ if (2.2) holds and the negative Fourier coefficients vanish when $\infty \in \overline{\tilde{U}}$. Given an open set $U \subset X(1)$, let $f \in \mathcal{O}_{X(1)}(U)$ be a modular form on the preimage $\pi^{-1}U \cap \mathbb{H}$ of weight 0. We can view f as a function on U, where the value at $z \in U - \{\infty\}$ is the value at any of the preimages, and the value at ∞ is the zeroth Fourier coefficient.

Proposition 2.4.1. *The ringed space $(X(1), \mathcal{O}_{X(1)})$ is a Riemann surface.*

Sketch. The key point is to show that any point $x \in X(1)$ has a neighbourhood D with a homeomorphism z, called a local coordinate or parameter, to a disk in $\mathbb{C}$, such that holomorphic functions on both disks coincide. There are three cases: $x = \infty$, x is an image of one of the fixed points $i, e^{2\pi i/3}$, or x is any other point. The first case was essentially done in the last section, q is the local coordinate at ∞. The third case is straight forward. The map $\pi : \mathbb{H} \to X(1)$ is unramified over x A local coordinate z at a point $y \in \mathbb{H}$ lying over x will give a local coordinate at x. The map π is ramified at i and $e^{2\pi i/3}$ with ramification index $e = 2$ and 3 respectively. z^e will give a local coordinate at the image. □

It is worth noting that the images of i and $e^{2\pi i/3}$ are nonsingular, and therefore no different from any other point from this point of view. However, these points clearly are special. One way to keep track of this, is the to use the language of orbifolds or stacks. To simplify our story, we won't do this here.

Given an integer $N > 0$, the *principal congruence subgroup of level N* of $\Gamma(1) = SL_2(\mathbb{Z})$ is

$$\Gamma(N) = \ker \Gamma(1) \to SL_2(\mathbb{Z}/N) = \{M \in \Gamma(1) \mid M \equiv I \mod N\}$$

A *congruence* group is a subgroup of $\Gamma(1)$ containing some $\Gamma(N)$. It therefore has finite index in $\Gamma(1)$. Some other important examples are

$$\Gamma_1(N) = \{M \in \Gamma(1) \mid M \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \mod N\}$$

$$\Gamma_0(N) = \{M \in \Gamma(1) \mid M \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \mod N\}$$

We have inclusions

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \Gamma(1)$$

We can compute the indices.

Proposition 2.4.2.

(a)

$$[\Gamma(1) : \Gamma(N)] = N^3 \prod \left(1 - \frac{1}{p^2}\right)$$

where p runs over primes dividing N.

(b)

$$[\Gamma(1) : \Gamma_1(N)] = N^2 \prod \left(1 - \frac{1}{p^2}\right)$$

(c)

$$[\Gamma(1) : \Gamma_0(N)] = N \prod \left(1 + \frac{1}{p}\right)$$

Proof. We have $[\Gamma(1) : \Gamma(N)] = |SL_2(\mathbb{Z}/N)|$, $[\Gamma_1(N) : \Gamma(N)| = |Z/N|$ and $[\Gamma_0(N), \Gamma_1(N)] = |(\mathbb{Z}/N)^*|$. These can be checked to yield the above formulas. □

Lemma 2.4.3. *$\Gamma(N)$ is torsion free once $N \geq 3$.*

Given such a group, it will act on $\mathbb{H}^*$, let $Y(\Gamma') = \Gamma' \backslash \mathbb{H}$ and let $X(\Gamma') = \Gamma' \backslash \mathbb{H}^*$. The points of $X(\Gamma') - Y(\Gamma')$ are called cusps. We write $Y(N), Y_1(N)$ etc. when the groups are $\Gamma(N), \Gamma_1(N)$. A meromorphic function f on $\mathbb{H}$ is weakly modular form of weight $2k$, with respect to Γ', if (2.2) holds for matrices in Γ'. The isotropy group of ∞ is a finite index subgroup of $\langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \rangle$ so it is of the form $\langle \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \rangle$ for some n. This implies that a weakly modular form satisfies

$$f(z + n) = f(z)$$

So that it has a Fourier expansion in $q = e^{2\pi iz/n}$. A similar Fourier expansion occurs at all the other cusps. We say that f is a modular (resp. cusp) form if it is holomorphic in $\mathbb{H}$ and the negative (resp. nonpositive) Fourier coefficients at each cusp vanish. When extend this to the case where the domain of f is an invariant open set. Then we can turn $Y(\Gamma') \subset X(\Gamma')$ into Riemann surfaces exactly as above. These are called *modular curves*. We have a holomorphic map

$$X(\Gamma') \to X(\Gamma(1)) = A_1 \cong \mathbb{P}^1$$

induced by inclusion $\Gamma' \subset \Gamma(1)$. This is a branched covering. So we can compute the genus using the Riemann-Hurwitz formula, which says that if $Y \to X$ is a degree d branched covering of compact Riemann surfaces of genus $g(Y)$ and $g(X)$, then

$$2g(Y) - 2 = (2g(X) - 2)d + \sum_{y \in Y} (e_y - 1)$$

where e_y is the ramification index which counts the number of sheets which "come together" at y. We will use this to compute for most of the principal congruence groups. More general formulas can be found in [DS, S2].

Theorem 2.4.4. *When $N \geq 3$, the genus of $X(N) = X(\Gamma(N))$ is*

$$g = 1 + \frac{d(N-6)}{12N}$$

where

$$d = \frac{1}{2}[\Gamma(1) : \Gamma(N)] = \frac{N^3}{2} \prod \left(1 - \frac{1}{p^2}\right)$$

The genus of $X(2)$ is zero.

Proof. The covering $\pi : X(N) \to X(1)$ is Galois with group $G = PSL_2(\mathbb{Z})/\operatorname{im}\Gamma(N) = PSL_2(\mathbb{Z}/N)$. The degree of this covering $|G| = d$, when $N \geq 3$, and $d = 6$ when $N = 2$. Let p_2 and p_3 represent the images of i and $e^{2\pi i/3}$ in $X(1)$. Then p_2, p_3, ∞ are the ramification points. Given one of these points p, and $q \in \pi^{-1}(p)$, e_q is the order of the isotropy group $G_q = \{g \in G\ gq = q\}$. This independent of q, because all the isotropy groups are conjugate. It also follows that $|\pi^{-1}(p)| = d/|G_q|$. So we can make a table consisting of p, $|\pi^{-1}(p)|$, $|e_q|$:

$$p_2, d/2, 2$$

$$p_3, d/3, 3$$

$$\infty, d/N, N$$

Putting these into Riemann-Hurwitz and simplifying proves the theorem. □

Using this formula, we can see that the first nonzero value for g occurs at $N = 7$, then $g = 3$. Note that $X(7)$ has an action of $PSL_2(\mathbb{Z}/7)$. The cardinality of this $168 = 84(g-1)$, which is the maximal possible size for an automorphism group by a theorem of Hurwitz. Formulas for the genera of other modular curves can be found in [DS, S2].

2.5 Dimension of spaces of modular forms

Given a smooth curve X and a divisor D, let $\Omega^1_X(D) = \Omega^1_X \otimes \mathcal{O}_X(D)$. It can be identified with $\mathcal{O}_X(K+D)$, where K is a canonical divisor. The space of global sections $\Gamma(X, \Omega^1_X(D))$ can be identified with the space of meromorphic 1-forms ω satisfying $\operatorname{div}\omega + D \geq 0$.

Theorem 2.5.1. *Suppose that Γ' is a torsion free congruence group. Let $X = X(\Gamma')$ and let $D = \sum p_i$ be the sum of cusps. The space weight $2k$ modular forms (resp. cusp forms) $M_{2k}(\Gamma')$ (resp. $S_{2k}(\Gamma')$) is isomorphic to $\Gamma(X, \mathcal{O}(kK + kD))$ (resp. $\Gamma(X, \mathcal{O}(kK + (k-1)D))$. In particular, $S_2(\Gamma') \cong \Gamma(X, \Omega^1_X)$*

Proof. Let $f(z) \in M_{2k}(\Gamma')$. Then $f(z)(dz)^{\otimes k}$ is a Γ'-invariant holomorphic section of $(\Omega^1_{\mathbb{H}})^{\otimes k}$, so it descends to a holomorphic section of $(\Omega^1_{Y(\Gamma')})^{\otimes k}$. We have to check what happens near a cusp. We have a local coordinate $q = e^{2\pi i z/n}$. By assumption f can be expanded as $\sum_0^\infty a_m q^m$, with $a_0 = 0$ for a cusp form. We have $dz = (n/2\pi i)dq/q$. So

$$f(z)(dz)^{\otimes k} = \left(\frac{n}{2\pi i}\right)^k (a_0 q^{-k} + a_1 q^{1-k} + \ldots)dq^{\otimes k}$$

So the theorem follows.

□

Corollary 2.5.2. *Suppose that X has genus g with m cusps, then*

$$\dim S_{2k}(\Gamma') = \begin{cases} g & \text{if } k = 1 \\ (2k-1)(g-1) + (k-1)m & \text{if } k > 1 \end{cases}$$

Proof. The first case is an immediate consequence of the theorem. For the second, we use Riemann-Roch.

$$\begin{aligned} h^0(\mathcal{O}(kK + (k-1)D)) &= h^0(\mathcal{O}(kK + (k-1)D)) - h^0(\mathcal{O}((1-k)K - (k-1)D) \\ &= \deg(kK + (k-1)D) + 1 - g \end{aligned}$$

□

A product of a modular form of weight $2k$ and 2ℓ is clearly a modular form of weight $2(k+\ell)$. Therefore $\bigoplus_k S_{2k}(\Gamma')$ is a graded $\mathbb{C}$-algebra.

Corollary 2.5.3. *The algebra of modular forms is finitely generated.*

Proof. This follows from the standard fact that the algebra

$$\bigoplus_k H^0(X, \mathcal{O}(kE))$$

is finitely generated, whenever X is a compact Riemann surface and E is divisor with $\deg E \geq 0$. □

We refer to [DS, S2] for more general formulas allowing k to be odd and Γ' to have torsion. Using these formulas, one can show that the algebra of modular forms for $SL_2(\mathbb{Z})$ is generated by the Eisenstein series G_4 and G_6.

2.6 Moduli interpretation

As we saw, $Y(1)$ parameterizes elliptic curves. While it's intuitively clear what this means, the actual statement requires a bit more precision. Let us define an analytic family of (compact) complex manifolds to be a (proper) holomorphic submersion of complex manifolds $f : E \to B$. We recall that a submersion is map such that derivative is surjective on tangent spaces. This implies that fibres $E_b = f^{-1}(b)$ are complex submanifolds. By an analytic family of elliptic curves we mean an analytic family of compact complex manifolds $f : E \to B$ with a holomorphic section $s : B \to E$ such that each fibre E_b is a compact Riemann surface of genus one. We can regard E_b as an elliptic curve with origin $s(b)$. Given an elliptic curve E, recall that we have a number $j(E) \in \mathbb{C}$ defined by an explicit formula.

Theorem 2.6.1. $Y(1)$ *has the following properties:*

(a) The map $E \mapsto j(E)$ *gives a bijection between the set of isomorphism classes of elliptic curves over* $\mathbb{C}$ *and points of* $Y(1)$*,*

(b) Given an analytic family elliptic curves $E \to B$*, the map* $B \to Y(1)$*, called the classifying map, given by* $b \mapsto j(E_b)$ *is holomorphic.*

The statement can be strengthened to completely characterize $Y(1)$, but we need a bit of terminology. Let $Ell^{an}(B)$ be the set of isomorphism classes of analytic families of elliptic curves over B, where isomorphism has the obvious meaning. Given a holomorphic map $B' \to B$, the pullback $E \mapsto E \times_B B'$ gives a map $Ell^{an}(B) \to Ell(B')$ which makes it into a contravariant functor. More generally, let $M(-)$ be contravariant functor from the category of complex manifolds (or schemes or...) to sets; one thinks of elements of $M(B)$ as families of objects over B of interest. We say that M is *representable* by U, or that U is a *fine moduli space* for M, if there is a natural isomorphism of functors

$$M(B) \cong Hom(B, U)$$

Yoneda's lemma, in category theory, tells us that U is completely determined by this property, and moreover it carries a universal family such that any object in $M(B)$ is the pullback of it under some map $B \to U$. Although this is ideal scenario for any moduli problem, it fails for Ell^{an}. This is because there exists nontrivial families in $Ell^{an}(B)$ with constant j-invariant. Here is a general construction.

Example 2.6.2. *Let* E *be either* E_i *or* $E_{\exp(2\pi i/3)}$ *Either curve has a nontrivial automorphism group* G*, which is cyclic in both cases. Choose a manifold* $\tilde{B}$ *on which* G *acts freely, e.g.* $\mathbb{C}^*$*. The quotient* $(E \times \tilde{B})/G \to \tilde{B}/G$ *is a nontrivial family with constant* j*-invariant.*

In spite of this bad news, we do have a natural transformation $Ell^{an}(B) \to Hom(B, Y(1))$, which is universal in an appropriate sense, and which induces a bijection when B is a point. We say that $Y(1)$ is the *coarse moduli space* for Ell^{an}.

The other modular curves have similar interpretations. Let us explain the characterizations of $Y(N) = Y(\Gamma(N))$ in a somewhat informal way. Given $E_\tau = \mathbb{C}/\mathbb{Z} + \mathbb{Z}\tau$, the image of $(\frac{1}{N}, \frac{\tau}{N})$ gives a basis for the N-torsion points in E_τ. We refer to this as a level N-structure; we will give a more precise definition below. If $A \in \Gamma(N)$, then the induced isomorphism $E_\tau \cong E_{A\cdot\tau}$ takes the above level N-structure of the first curve to the level structure of the second. In order to make this notion independent of our representation of E_τ as a quotient, we note that the lattice is isomorphic to homology $L_\tau \cong H_1(E_\tau, \mathbb{Z})$. Thus a level N-structure is a choice of basis for $H_1(E, \mathbb{Z}/N\mathbb{Z}) = H_1(E, \mathbb{Z}) \otimes \mathbb{Z}/N\mathbb{Z}$, but not just any basis. The group carries an intersection pairing

$$H_1(E, \mathbb{Z}/N) \times H_1(E, \mathbb{Z}/N\mathbb{Z}) \to \mathbb{Z}/N$$

So now we can give the precise definition. A level N-structure is a basis for $H_1(E, \mathbb{Z}/N\mathbb{Z})$, which is symplectic in the sense that the matrix of the above pairing is

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

Theorem 2.6.3. *$Y(N)$ is the coarse moduli space of elliptic curves with level N-structure. When $N \geq 3$ it is a fine moduli space.*

Recall that the assumption $N \geq 3$ is precisely the condition to guarantee that $\Gamma(N)$ is torsion free. This same condition also allows us to kill the automorphism groups which created the problem in example 2.6.2.

For the other moduli spaces, we have similar interpretations. $Y_1(N) = Y(\Gamma_1(N))$ is the coarse moduli space of pairs (E, P) consisting of an elliptic curve E and a point P of order N. $Y_0(N)$ is the moduli space of pairs (E, C) consisting of an elliptic curve and a cyclic subgroup of the group of N-torsion points. The projections

$$Y(N) \to Y_1(N) \to Y_0(N) \to Y(1)$$

induced by the inclusions

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \Gamma$$

have moduli interpretations. Given E a level N-structure is a pair of N-torsion points P, Q satisfying suitable conditions. The map $Y(N) \to Y_1(N)$ corresponds to the forgetful map $(E, P, Q) \mapsto (E, P)$.

2.7 Models over number fields

So far we have considered modular curves as Riemann surfaces, but in fact they are algebraic curves. This is true of any compact Riemann surface minus a finite set of points. However, a more natural way to see this is to consider algebraic versions of the moduli problems consider earlier. As a bonus this will show that these curves are naturally defined over number fields and even over rings of integers. This is very important for applications to number theory. Let us start with $Y(1)$. We consider the corresponding moduli problem in the algebraic setting. Given a scheme B, an elliptic curve over it is a smooth proper map is a smooth proper map $f : E \to B$, with a section, such that the closed fibres of f are genus one curves. Let $Ell(B)$ denote the isomorphism classes of elliptic curves over B. Then $Y(1)_{\mathbb{Z}} = \operatorname{Spec} \mathbb{Z}[j]$ is a coarse moduli scheme for $Ell(-)$, and this gives a model for $Y(1)$ over the integers, i.e. $Y(1)$ is the complex manifold associated to $Y(1)_{\mathbb{Z}} \times_{\operatorname{Spec} \mathbb{Z}} \operatorname{Spec} \mathbb{C}$.

Next let us turn to $Y(N)$. We can formulate the definition of level structure of an elliptic curve over an arbitrary field k. In this case, we need N to be prime to the characteristic. Then a level N-structure is a pair of N-torsion points $P, Q \in E(k)$ such that they generate the group of N-torsion points and such that $e_N(P, Q)$ is a primitive N-root of unity. Here e_N is the Weil pairing whose definition can be found in [Si]. Note that the condition forces k to contain a primitive N-root of unity. More generally, there is a notion of a level structure for an elliptic curve over a base scheme. This is basically a pair of sections which induces a level structure on the closed fibres.

Theorem 2.7.1. *There exists a scheme $Y(N)$ defined over $\mathbb{Z}[1/N, e^{2\pi i/N}]$ which is the coarse moduli space of elliptic curves with level N-structure. It is a fine moduli when $N \geq 3$. The set of complex points is the Riemann surface $\Gamma(N)\backslash\mathbb{H}$ considered before.*

See Deligne-Rapoport [DR] for the construction in general. They also give a more general construction which would include the $Y_i(N)$. $Y_0(N)$ is particularly interesting because it is defined over $\mathbb{Q}$. When N is small, $Y(N)$ can be made very explicit. We have

$$Y(2) = \operatorname{Spec} \mathbb{Z}[\frac{1}{2}, t, \frac{1}{t(t-1)}]$$

Although this is not fine, there is an "almost" universal family called the Legendre family

$$y^2 z = x(x - z)(x - tz)$$

in $\mathbb{P}^2_{\mathbb{Z}}$. Over a field, this curve has 4 branch points over $0, 1, t, \infty$. Take the first to be the origin, and the next two to be the level 2-structure.

When $N = 3$, let $R = \mathbb{Z}[1/3, e^{2\pi i/3}]$, then

$$Y(N) = \operatorname{Spec} R[t, \frac{1}{t^3 - 1}]$$

The universal family is given by the elliptic curve

$$x^3 + y^3 + z^3 = 3txyz$$

in $\mathbb{P}^2_R$ with section $[1, -1, 0]$. The level 3-structure is given by the sections $[-1, 0, 1]$ and $[-1, e^{2\pi i/3}, 0]$.

Chapter 3

Hilbert and Picard modular surfaces

3.1 The Hilbert modular group

Let $D > 0$ be square free integer (D is not divisible by any squares other than 1). Let $K = \mathbb{Q}(\sqrt{D})$ be the corresponding real quadratic field. This is Galois over $\mathbb{Q}$, with Galois group generated by the involution $(a + b\sqrt{D})' = a - b\sqrt{D}$. The norm $N(x) = xx'$ and trace $\mathrm{tr}(x) = x + x'$. There are two embeddings of fields $\sigma_i : K \hookrightarrow \mathbb{R}$ given by

$$\sigma_1(a + b\sqrt{D}) = a + b\sqrt{D}$$
$$\sigma_2(a + b\sqrt{D}) = a - b\sqrt{D}$$

This gives an embedding of groups $SL_2(K) \hookrightarrow SL_2(\mathbb{R})^2$ by $A \mapsto (\sigma_1(A), \sigma_2(A))$.

The ring of integers $\mathcal{O}_K \subset K$ is the integral closure of $\mathbb{Z}$ in K. More explicitly,

$$\mathcal{O}_K = \begin{cases} \mathbb{Z} + \frac{1+\sqrt{D}}{2}\mathbb{Z} & \text{if } D \equiv 1 \mod 4 \\ \mathbb{Z} + \sqrt{D}\mathbb{Z} & \text{if } D \equiv 2, 3 \mod 4 \end{cases}$$

$\mathcal{O}_K$ is a Dedekind domain, so we can define the class group $Cl(\mathcal{O}_K)$ in the usual way, as the group of fractional ideals modulo principal ideals. This is a finite group; its cardinality h is called the class number of K.

Lemma 3.1.1. *If we embed $\mathcal{O}_K \hookrightarrow \mathbb{R}^2$ by $\sigma_1 \times \sigma_2$, then the image is discrete.*

This is false with only a single σ_i.

Proof. This is elementary. For example, when $D \equiv 2, 3 \mod 4$, this follows from the inequality

$$(a + b\sqrt{D})^2 + (a - b\sqrt{D})^2 = a^2 + b^2 D \geq 1$$

for a nonzero integer $a + b\sqrt{D}$. □

The *Hilbert modular group* (for a given K) is $\Gamma_K = SL_2(\mathcal{O}_K)$. This embeds into $SL_2(\mathbb{R})^2$ as above. From the previous lemma, we easily deduce:

Corollary 3.1.2. *The image of the Hilbert modular group in $SL_2(\mathbb{R})^2$ is discrete with respect to the usual topology.*

Given a nonzero ideal $I \subset \mathcal{O}_K$, we can define the corresponding principal congruence group as

$$\Gamma_K(I) = \ker[SL_2(\mathcal{O}_K) \to SL_2(\mathcal{O}_K/I)]$$

This applies, in particular, to an ideal of the form (N), N where N is a nonzero integer.

Lemma 3.1.3. *$\Gamma_K(N)$ is torsion free when $N \geq 3$.*

Proof. [F, p 42]. □

We will refer to a subgroup containing some $\Gamma(N)$ as a conguence group.

3.2 Hilbert modular surfaces: topology

Fix $K = \mathbb{Q}(\sqrt{D})$ as before. The group Γ_K acts on $\mathbb{H}^2$ through its embedding into $SL_2(\mathbb{R})^2$. To be more explicit, given $(z_1, z_2) \in \mathbb{H}^2$ and $A \in \Gamma_K$, $A \cdot (z_1, z_2) = (A \cdot z_1, A \cdot z_2)$. The action factors through $PSL_2(\mathcal{O}_K) = SL_2(\mathcal{O}_K)/\{\pm I\}$.

Proposition 3.2.1. *This action is properly discontinuous.*

Proof. This follows from the discreteness of Γ_K by [F, p 21].

□

Let $\Gamma \subseteq \Gamma_K$ be a congruence subgroup. The quotient $X^o(\Gamma) = \Gamma\backslash\mathbb{H}^2$, and various related objects, are called *Hilbert modular surfaces* or sometimes Hilbert-Blumenthal surfaces. The proposition implies that the quotient topology has reasonable properties (e.g. it is Hausdorff). However, the action of $PSL_2(\mathcal{O}_K)$ is not free. For instance, $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ fixes (i, i). If we restrict to a torsion free subgroup such as $\Gamma_K(3)$, then the action becomes free. It follows that isotropy subgroup of Γ for any point of $\mathbb{H}^2$ is finite. In fact, one can show that they are cyclic. We claim that the image of any fixed point in $X^o(\Gamma)$ is topologically singular in the sense that $X^o(\Gamma)$ is not even a topological manifold at that point. To see this, we can work locally. A local model for this is given as follows. Choose a finite cyclic subgroup $G \subset GL_2(\mathbb{C})$, such that $\mathbb{C}^2$ has no nonzero invariant vectors. The quotient $\mathbb{C}^2/G$ is a singular algebraic variety. In the simplest example, $G = \{\pm 1\}$, $u = x^2, v = y^2, w = xy$ generate the ring of invariant polynomials, and $\mathbb{C}^2/G = V(uv - w^2)$. A n-manifold X has the property that any point $x \in X$ has a fundamental system of neighbourhoods U, such that that $U - x$ is homotopy equivalent to S^{n-1}. This is not true for $\mathbb{C}^2/G$, with its usual topology.

Finally, we note that $X^o(\Gamma)$ is not compact. We can compactify it, as we did for modular curves, by adding cusps. We embed $\mathbb{P}^1(K) \to \mathbb{P}(\mathbb{R})^2$ by $\sigma_1 \times \sigma_2$. Using this, we can regard points of $\mathbb{P}(K)$ as lying on the boundary of $\mathbb{H}^2$ via

$$\mathbb{P}^1(\mathbb{R})^2 \subset \mathbb{P}^1(\mathbb{C})^2 \supset \mathbb{H}^2$$

A Γ-orbit of point of $\mathbb{P}^1(K)$ is called a *cusp* with respect to Γ. Given a point $[a,b] \in \mathbb{P}^1(K)$, we let (a,b) denote the fractional ideal generated by these elements. Although this ideal is not well defined, its class in $Cl(\mathcal{O}_K)$ is.

Proposition 3.2.2. *The above map gives a bijection between the set of cusps for* Γ_K *and* $Cl(\mathcal{O}_K)$.

Proof. Let $\phi : \mathbb{P}^1(K) \to Cl(\mathcal{O}_K)$ denote the above map. It is known that any fractional ideal of $\mathcal{O}_K$ is generated by two elements. Therefore ϕ is surjective. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then

$$\phi(A \cdot [x,y]) = (ax + by, cx + dy) \subseteq (x,y)$$

The same argument, using A^{-1}, gives the opposite inclusion. Therefore ϕ factors through a map $\bar{\phi} : \Gamma_K \backslash \mathbb{P}^1(K) \to CL(\mathcal{O}_K)$.

It remains to prove that $\bar{\phi}$ is injective. We assume two points of $\mathbb{P}^1(K)$ have the same image under ϕ. For simplicity, we treat the case where on the points is $\infty = [1,0]$. Denote the other by $[x,y]$. We can assume that both $x, y \in \mathcal{O}_K$. Since $(x,y) = \phi(\infty) = (1)$, we must have $ax + by = 1$ for some $a, b \in \mathcal{O}_K$. Then

$$A = \begin{pmatrix} a & b \\ -y & x \end{pmatrix}$$

lies Γ_K and it maps $[x,y]$ to ∞. Therefore they lie in the same Γ_K-orbit. □

Corollary 3.2.3. Γ_K *has* h *cusps. A congruence subgroup* $\Gamma \subset \Gamma_K$ *has a finite number of cusps.*

Let $(\mathbb{H}^2)^* = \mathbb{H}^2 \cup \mathbb{P}^1(K)$. We put a topology on this, such that

1. It agrees with the usual one on $\mathbb{H}^2$
2. The sets of form

$$U_C = \{(z_1, z_2) \mid \mathrm{Im}\,(z_1)\mathrm{Im}\,(z_2) > C\} \cup \{\infty\}, \quad C \in \mathbb{R}^+$$

   forms a fundamental systems of neighbourhoods of ∞.
3. If $p = A\infty$, with $A \in SL_2(K)$, then AU_C forms a fundamental system of neighbourhoods of p.

Let $\Gamma \subset \Gamma_K$ be a congruence group. Let $X(\Gamma) = \Gamma \backslash (\mathbb{H}^2)^* = X^o(\Gamma) \cup \{\text{cusps}\}$ with the quotient topology. To analyze the quotient, we let Γ_p be the isotropy group of a cusp p. The structure of this group is essentially given as follows.

Lemma 3.2.4. *There exists a rank 2 group additive subgroup $M \subset K$, and finite index multiplicative subgroup $V \subseteq \{u \in \mathcal{O}_K^* \mid u$ totally positive, i.e. $\sigma_i(u) > 0\}$ such that the V stabilizes M, and the group*

$$G(M,V) = \{\begin{pmatrix} \epsilon & \mu \\ 0 & \epsilon^{-1} \end{pmatrix} \mid \mu \in M, \epsilon \in V\}$$

has finite index in Γ_p.

Lemma 3.2.5.

1. *For any $C > 0$, Γ_∞ stabilizes U_C and $\Gamma_\infty \backslash \bar{U}_C$ is compact.*
2. *For $C \gg 0$, the image of U_C in $X(\Gamma)$ is homeomorphic to $\Gamma_\infty \backslash U_C$*
3. *If $p = A\infty$ is a cusp different from ∞, i.e. if p does not lie in the Γ orbit of ∞, then for $C \gg 0$, the images of U_C and AU_C in $X(\Gamma)$ are disjoint.*

Proof. See [G2, pp 7-9]. □

Theorem 3.2.6. *$X(\Gamma)$ is compact Hausdorff.*

Proof. Using the previous lemma, one can build a compact fundamental domain. See [F, p 38] for details. □

3.3 Hilbert modular forms

In order to discuss the analytic properties of Hilbert modular surfaces, we need the appropriate category. An analytic space is to a complex manifold what an algebraic variety, or more generally scheme, is to a nonsingular variety. A basic example is to start with an open ball $B \subset \mathbb{C}^N$, choose a collection of holomorphic functions $f_1, f_2, \ldots, f_n$ and consider the zero set $Z = \{x \in B \mid f_i(x) = 0\}$. We refer to this as a model. In general, an analytic space is something which locally looks like a *model*. To make this more precise, we can proceed as with scheme theory (for those familiar with them) by introducing a sheaf. For our model Z, let $\mathcal{O}_Z$ denote the sheaf of restrictions of holomorphic functions from B to Z. A (reduced) *analytic space* is a pair $(X, \mathcal{O}_X)$ consisting of a paracompact Hausdorff space, and a sheaf of continuous complex valued functions, such that it is locally isomorphic (as a locally ringed space) to a pair given by a model. See Grauert-Remmert [GR] for further details (where analytic spaces are called complex spaces). An analytic space $(X, \mathcal{O}_X)$ is called *normal* if all its stalks $\mathcal{O}_{X,x}$ are integrally closed. Here are a couple of examples.

1. Complex manifolds are normal because the stalks are rings of convergent power series, and these are regular noetherian and therefore integrally closed.
2. If $G \subset GL_n(\mathbb{C})$, then $\mathbb{C}^n/G$ is normal. Away from 0, it's a manifold, The stalk at 0 is the ring of G-invariant convergent power series, and this is easily seen to be integrally closed.

The importance of this condition for us stems from the following. Suppose that X is normal, then:

1. The set of singular points (points where it fails to be a manifold) has codimension at least 2 [GR, p 128]

2. Any holomorphic function defined on the nonsingular part of X extends to a holomorphic function on X [GR, p 144]

Here is a useful criterion.

Theorem 3.3.1 (Cartan)**.** *Suppose that $X - \{x\}$ is a normal analytic space, and there is a system of neigbourhoods U of x such that $U - \{x\}$ is connected and such that holomorphic functions in $U - \{x\}$ separate points. If we define $f : U \to \mathbb{C}$ to be holomorphic if it is continuous and holomorphic away from x, then X becomes a normal analytic space.*

Proof. [C, exp 11]. □

Let K, $(\mathbb{H}^2)^*, \Gamma \subseteq \Gamma_K$ and $X = X(\Gamma)$ be as before. We have projections $\pi : (\mathbb{H}^2)^* \to X$ and $\pi^o : \mathbb{H}^2 \to X$. If $U \subseteq X$ is open, define $f \in \mathcal{O}_X(U)$ if it continuous and if $f \circ \pi^o$ is holomorphic.

Theorem 3.3.2. *$(X, \mathcal{O}_X)$ is a normal analytic space.*

Proof. Away from the cusps, this is easy by the above remarks. At a cusp, say ∞, one checks Cartan's criterion holds. The topological condition for theorem 3.3.1 is clearly satisfied. We just have to check the second. Given a bounded holomorphic function f on U_C, the associated Poincaré series, which is the sum

$$\sum_{A \in \Gamma_\infty} f(A \cdot z)$$

can be shown to converge to a Γ_∞-invariant holomorphic function [F, p 57-58, p 113]. This construction yields sufficiently many holomorphic functions on $\Gamma_p \backslash U_{\mathbb{C}}$ to separate points. □

The singular points of X consist of images of the fixed points in $\mathbb{H}^2$ and the cusps. Since X is normal, we can define holomorphic functions, and related things, by prescribing them away from these points. So for example, a holomorphic function on X is given by a Γ_K-invariant function on $\mathbb{H}^2$. Such a function is necessarily constant by compactness of X. To get something more interesting, we have to relax the invariance condition. A Hilbert modular form of weight (k, ℓ), with respect to Γ, is a holomorphic function on $\mathbb{H}^2$ satisfying

$$f\left(\frac{az_1 + b}{cz_1 + d}, \frac{a'z_2 + b'}{c'z_2 + d'}\right) = (cz_1 + b)^k (c'z_2 + b')^\ell f(z_1, z_2) \tag{3.1}$$

for every element of Γ. We say this has weight k, when $k = \ell$, and this what we mostly care about. By a calculation similar to what we did earlier, we can see that f is a weight $2k$ modular form precisely when the tensor $f(dz_1 \wedge dz_2)^{\otimes k}$ is invariant under Γ_K. Note that, unlike the one dimensional case, we don't have to impose any extra holomorphicity conditions at the cusps, since this comes for free by normality. This fact, which can be checked to directly, goes by name of the "Koecher principle". To be see this, recall that is the cusp $p \in \mathbb{P}^1(K)$ is stable under translations by

$$G(M,V) \supset \left\{ \begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix} \mid \mu \in M \right\} \cong \mathbb{Z}^2$$

for some choice of (M, V). The functional equation (3.1) forces periodicity with respect to this group. Therefore we have a Fourier expansion. The extension property amounts to showing that that the negative degree Fourier coefficients automatically vanish. In fact, a bit more is true.

Proposition 3.3.3. *A holomorphic function f at p has a Fourier expansion*

$$f(z_1, z_2) = a_0 + \sum_{\nu \in M^\vee} a_\nu \exp(2\pi i(\nu z_1 + \nu' z_2))$$

where $M^\vee = \{\nu \in K \mid \forall \mu \in M, \operatorname{tr}(\mu\nu) \in \mathbb{Z}\}$. Furthermore, we have

$$a_{\epsilon\nu} = a_\nu,\ \forall \epsilon \in V \tag{3.2}$$

and $a_\nu = 0$ unless ν is 0 or totally positive.

Proof. The Fourier expansion and (3.2) follows from the invariance under the group $G(M, V)$. Suppose that $a_\nu \neq 0$ where $\nu \neq 0$ is not totally positive. Then for $\epsilon \in V$, $\epsilon > 1$, one finds that the sequence $\exp(-2\pi \operatorname{tr}(\epsilon^n \nu))$ is bounded away from 0. This, together with (3.2), would force the Fourier series to diverge along the ray $\{(ir, ir) \mid r \in \mathbb{R}^+\}$.

□

A modular form is called a *cusp form* if it vanishes at all the cusps, or equivalently if the zeroth Fourier coefficients are all zero. We will mainly be concerned with the case $\ell = k$, in which case we call this is a modular form of weight k. We denote the space of these by $M_k(\Gamma)$, and subspace of cusp forms by $S_k(\Gamma)$. We give a basic example. We try to form the series

$$\sum \frac{1}{[(cz_1 + d)(c'z_2 + d')]^k}$$

where k is even and (c, d) runs over $\mathcal{O}_K^2$. This will have the right formal properties, but it will diverge. The problem is that terms are repeated infinitely often. We can correct the problem by choosing (c, d) to range over a set of representatives for the orbit space $\mathcal{O}_K^2/\mathcal{O}_K^*$ under $(c, d) \mapsto (\epsilon c, \epsilon d)$. Then for $k > 2$, this will converge to an element of $M_k(\Gamma_K)$ called an Eisenstein series.

One of the things we can use Hilbert modular forms for is to embed X into projective space. If $f_0, \ldots, f_N$ is a collection of Hilbert modular forms of the same weight k, then we get a map $X \dashrightarrow \mathbb{P}^N$ by $x \to [f_0(x), \ldots, f_N(x)]$.

Theorem 3.3.4 (Baily-Borel)**.** *There are sufficiently many modular forms of some weight to get an embedding* $X \hookrightarrow \mathbb{P}^N$

Baily and Borel's [BB] theorem holds not just for Hilbert modular surfaces, but more generally for quotients of hermitian symmetric spaces by arithmetic groups. In particular, their theorem will apply to various other the examples considered later on. By using an extension of Chow's theorem due to Serre [GAGA], we obtain

Corollary 3.3.5. *X is a normal projective variety.*

This has a moduli interpretation, but the explanation will have to wait until we get to abelian varieties.

3.4 Riemann-Roch for surfaces

Fix $K, \ldots, X = X(\Gamma)$ as above. Since X is a normal projective surface, we can use methods from algebraic geometry to study it. We also need to appeal to Serre's GAGA theorem [GAGA], to switch from holomorphic to algebraic objects. As a first step, we need to resolve the singularities.

Theorem 3.4.1. *There exists a regular map $\pi : Y \to X$ such that*

(a) Y is nonsingular.

(b) π is an isomorphism over the nonsingular locus of X.

(c) If Y' is a nonsingular surface through which π factors, then $Y = Y'$

We fix one such surface $Y(\Gamma) = Y$, which is called a *minimal resolution* of X. We note that is unique under some additional assumptions, and it always exists by the general theory [BPV]. In the present case, it can be constructed quite explicitly [G2]. The explicit construction yields more information, which is needed to prove some of the results below. For example, it known that the preimage of a cusp (resp. non-cusp singularity) is a cycle (resp. chain) of rational curves.

We recall some basic facts from algebraic surface theory. Fix a nonsingular projective surface S over an algebraically closed field k. We will only need the case where $k = \mathbb{C}$, but state the results more generally when possible. A divisor on S is a finite sum $D = \sum n_i C_i$, where $n_i \in \mathbb{Z}$ and $C_i \subset X$ are possibly singular irreducible closed curves. Any such curve determines a discrete valuation ord_C on the field of rational functions $k(S)$, which measures the order of zero or pole along it. If f is a nonzero rational function on S, we can define the associated principal divisor $\operatorname{div} f = \sum ord_C(f)C$. The divisor class group $Cl(S)$ is defined as for curves by the abelian group of all divisors by the subgroup of principal divisors. If ω is a rational 2-form, we can define $\operatorname{div} \omega = \sum ord_C(\omega)C$ with a suitable definition of $ord_C(\omega)$. The divisor class is independent of ω and it is called the canonical divisor class K_S, or simply K (it is unlikely to be confused with the field). Given a divisor D, we define sheaf $\mathcal{O}_S(D)$ exactly as for curves. In particular, the space of global sections

$$H^0(S, \mathcal{O}_S(D)) = \{f \in k(S)^* \mid \operatorname{div} f + D \geq 0\} \cup \{0\}$$

This is finite dimensional, and we denote its dimension by $h^0(\mathcal{O}_S(D))$. The isomorphism class of $\mathcal{O}_S(D)$ depends only only the class of D in $Cl(S)$. The sheaf $\mathcal{O}_S(K)$ is isomorphic to the sheaf of regular 2-forms $\Omega^2_S = \wedge^2\Omega^1_S$. In particular, $h^0(\mathcal{O}(K)$ is the dimension of the space of regular 2-forms. This is one of the fundamental invariants of S, called the *geometric genus* $p_g(S)$. There is a new phenomenon for surfaces, namely divisor can be intersected. Suppose that C and D are distinct irreducible curves. Then $C \cap D$ is finite. If $p \in C \cap D$, define the intersection multiplicity at p by

$$(C \cdot D)_p = \dim \mathcal{O}_{S,p}/(f, g) = \dim \hat{\mathcal{O}}_{S,p}/(f, g)$$

where $\mathcal{O}_{S,p}$ is the local ring of the surface at p, and f, g are local equations of C and D in this ring. For example, this number is 1 if C and D are nonsingular and meet transervely at p, because f and g generate the maximal ideal of the completion $\hat{\mathcal{O}}_{S,p}$. Define the intersection number

$$C \cdot D = \sum_{p \in C \cap D} (C \cdot D)_p \tag{3.3}$$

Theorem 3.4.2. *There exists a bilinear form $Cl(S) \times Cl(S) \to \mathbb{Z}$ which agrees with the above intersection number whenever it is defined.*

Proof. See [H, chap 5, sec. 1]. □

We come to the main point. One would like a formula for $h^0(\mathcal{O}_S(D))$ for any divisor, but what one has is a formula for the Euler characteristic

$$\chi(\mathcal{O}(D)) = h^0(\mathcal{O}(D)) - h^1(\mathcal{O}(D)) + h^2(\mathcal{O}(D))$$

where h^i represent dimensions of higher cohomology groups. One can view the higher h^i above as corrections. These will be dealt with shortly.

Theorem 3.4.3 (Riemann-Roch for surfaces)**.** *For any divisor*

$$\chi(\mathcal{O}(D)) = \frac{1}{2} D \cdot (D - K) + \chi(\mathcal{O}_S)$$

When $k = \mathbb{C}$

$$\chi(\mathcal{O}_S) = \frac{K^2 + e(S)}{12}$$

where $e(S)$ is the topological Euler characteristic.

Proof. For the first formula, see [H, chap 5, sec. 1]. The second is special case of the Hirzebruch-Riemann-Roch theorem [BPV, p 20] □

In order to get an exact formula for $h^0(\mathcal{O}(D))$, we need to combine Riemann-Roch with a so called vanishing theorem. We use the famous Kodaira vanishing theorem. We work over $\mathbb{C}$ since the result can fail in positive characteristic. A divisor D is called ample if $\mathcal{O}(nD))$ has sufficiently many sections to give an embedding of S into projective space for all $n \gg 0$.

Theorem 3.4.4 (Kodaira Vanishing). *Assume $k = \mathbb{C}$. If D is an ample divisor, then*

$$H^i(X, \mathcal{O}(K + D)) = 0 \quad i > 0$$

Proof. See [GH, chap 1] for a proof. We remark that it true in arbitrary dimension. Kodaira's original formulation, which is used in the reference, is that the line bundle $\mathcal{O}(D)$ carries metric with positive curvature. This is equivalent to ampleness by Kodaira's embedding theorem. There is now a purely algebraic proof, due to Deligne and Illusie, which uses the above formulation □

Let us return to our Hilbert modular surface Y. We assume for simplicity that Γ is torsion free. The union of preimages of the cusps in X forms a divisor D. The structure of D can be determined rather explicity [G2, chap 2]. As we saw, an element of $M_2(\Gamma)$ gives an invariant holomorphic 2-form on $\mathbb{H}^2$ and therefore a holomorphic form on $X^o(\Gamma)$. If it vanishes at the cusps, then it would extend to a holomorphic form, and therefore regular form, on Y. In fact, the converse holds also.

Theorem 3.4.5. *There is an isomorphism between $S_2(\Gamma) \cong H^0(\mathcal{O}_Y(K))$. Therefore the dimension of this space is the geometric genus $p_g(Y)$.*

Proof. [G2, p 57]. □

For higher weight, we have:

Theorem 3.4.6. *The divisor $F = K + D$ is ample, and for any $m > 0$, we have an isomorphism*

$$S_{2m}(\Gamma) \cong H^0(Y, \mathcal{O}_Y(K + (m-1)F))$$

Proof. [G2, p 72]. □

From Kodaira vanishing plus Riemann-Roch, we obtain:

Corollary 3.4.7.

$$\dim S_{2m}(\Gamma) = \frac{m-1}{2}(K + (m-1)F) \cdot F + \chi(\mathcal{O}_Y)$$

The right side can be evaluated in explicit terms. See [G2, chap IV, sect 4] for a more complete discussion.

3.5 Picard modular surfaces

Let

$$B = \{(z_1, z_2) \in \mathbb{C}^2 \mid |z_1|^2 + |z_2|^2 < 1\}$$

denote the ball in $\mathbb{C}^2$. This can be identified with the subset

$$B' = \{[z_1, z_2, z_3] \in \mathbb{P}^2 \mid |z_1^2 + |z_2|^2 - |z_3|^2 < 0\}$$

of the complex projective plane, under $(z_1, z_2) \mapsto [z_1, z_2, 1]$. If we introduce the nonpositive definite hermitian form H on $\mathbb{C}^3$ with matrix

$$J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

then

$$B' \cong \{[v] \in \mathbb{P}^2 \mid H(v, v) < 0\}$$

Let

$$SU(2, 1) = \{A \in GL_3(\mathbb{C}) \mid \bar{A}^T JA = J\,, \det A = 1\},$$

denote the special unitary group associated to H. Clearly this group acts on B' The following is straightforward.

Lemma 3.5.1. *$SU(2, 1)$ acts transitively on B'.*

Now we fix a square free integer $D > 0$ and consider the imaginary quadratic field $K = \mathbb{Q}(\sqrt{-D})$. We have embeddings $\sigma_i : K \to \mathbb{C}$, and involution $x \mapsto x'$, and the ring of integers $\mathcal{O}_K$ described exactly as in the real quadratic case. Let $V = K^3$ with a lattice $L \subset V$, i.e. an $\mathcal{O}_K$-submodule such that $K \otimes L \cong V$. We assume that V is equipped with a form H_0 which is Hermitian in the sense that $H_0(ax, y) = aH_0(y, x)$, $H_0(x, y) = H_0(y, x)'$ for $a \in K$ and $x, y \in V$, and $\mathcal{O}_K$-valued on L. We suppose that after extending scalars to $\mathbb{C}$, using either σ_1 or σ_2 (it won't matter), H_0 has signature $(2, 1)$. For example, $H_0(x, y) = x^T Jy'$ satisfies these conditions, but it is not the only choice. We can form the special unitary group $SU(H_0)$, which is the subgroup of $SL_3(K)$ of linear transformations which preserve H_0. The subgroup $\Gamma_{K,H_0} \subset SU(H_0)$ stabilizing the lattice L is called a *Picard modular group.* We will fix the above data, and simply denote this group by Γ. By our assumptions, H_0 can be identified with H after extending scalars, so we can embed $\Gamma \subset SU(2, 1)$. This gives an action of Γ on B. The action is seen to be properly discontinuous. Therefore the quotient $X^o = X^o(\Gamma, H_0) = \Gamma\backslash B$ inherits a Hausdorff topology.

The space X^o is not compact. We can compactify it by adding a finite number of points, again called cusps. Using the model B', we see that it has a natural boundary $\partial B'$ consisting of lines generated by vectors which isotropic in the sense that $H(v, v) = 0$. Let

$$B^* = B' \cup \{[v] \mid v \in \mathbb{Q}^3, v \text{ isotropic}\}$$

It is not difficult to see that Γ acts on this. Once again we have

Theorem 3.5.2 (Baily-Borel)**.** *For a suitable topology on B^*, the quotient $X = \Gamma \backslash B^*$ is compact. This can be given the structure of a normal algebraic surface.*

As above, we can construct a minimal resolution Y of X. This, and related objects, are called *Picard modular surfaces.*

Chapter 4

Abelian varieties

4.1 Abelian varieties

An abelian variety is a higher dimensional version of an elliptic curve. Here is the precise definition. Over a field k, an abelian variety consists of a smooth projective variety X over k, a k-rational point 0 and morphisms $+ : X \times X \to X$ and $- : X \to X$ which make it into a group. We have the following basic facts:

Theorem 4.1.1. *An abelian variety is a commutative group. When $k = \mathbb{C}$, an abelian variety has the structure of a complex torus, i.e. as a complex Lie group, it is isomorphic to $\mathbb{C}^g$ modulo a lattice.*

Proof. See Mumford [MAV, p 2, p 44]. □

We now focus on the case where $k = \mathbb{C}$. We can ask when is a complex torus $\mathbb{C}^g/L$ an abelian variety? The naive guess is that it is always true, but turns out to be incorrect once $g > 1$. A necessary condition for a compact complex manifold to be projective is that it has at least one nonconstant meromorphic function, but this can fail for higher dimensional tori. To formulate sufficient conditions, we modify what we did before with elliptic curves, but now we replace the element τ in the upper half plane with a $g \times g$ symmetric matrix Ω with positive definite imaginary part. The set of such matrices forms a complex manifold $\mathbb{H}_g$, that we call the Siegel upper half plane. Given such a matrix, we can form the lattice $L_\Omega = \mathbb{Z}^g + \Omega\mathbb{Z}^g$; $\Omega\mathbb{Z}^g$ means the group of integer linear combinations of columns of Ω. Define the complex torus $A_\Omega = \mathbb{C}^g/L_\Omega$. Since $\operatorname{Im}\Omega$ is positive definite, the terms in the series

$$\theta(z) = \sum_{n\in\mathbb{Z}^g} \exp(\pi i n^T \Omega n + 2\pi i n^T z)$$

go to zero rapidly as $||n|| \to \infty$. So it converges to a holomorphic function called the *Riemann theta function*. This is a generalization of the Jacobi function. It satisfies a similar functional equation

$$\begin{aligned}\theta(z+n) &= \theta(z)\\ \theta(z+\Omega n) &= \exp(-\pi i k n^t \Omega n - 2\pi i k z^t n)\theta(z)\end{aligned} \tag{4.1}$$

for all $n \in \mathbb{Z}^g$ and $k = 1$. More generally, a holomorphic function satisfying these equations is called a theta function of weight k. Any theta function can be expanded in a Fourier series by the first equation of (4.1), and the second leads to recurrence relations on the Fourier coefficients, which shows that the function is determined by finitely many coefficients. In particular, the space of theta functions of given weight can be seen to be finite dimensional.

Theorem 4.1.2 (Lefschetz)**.** *A_Ω is an abelian variety.*

Idea. Consider the vector space V of all theta functions of weight 3. If $f_0 \ldots, f_N$ is a basis of V, we can see that the point $[f_0(z), \ldots, f_N(z)] \in \mathbb{P}^N$ depends only on $z \mod L_\Omega$. We claim that

$$x \mapsto [f_0(z), \ldots, f_N(z)]$$

is defined everywhere and gives an embedding of $\phi : A_\Omega \to \mathbb{P}^N$. By considering products of the form

$$\theta(z+u)\theta(z+v)\theta(z-u-v) \in V, \quad u, v \in \mathbb{C}^g$$

we find a theta function which is nonzero at any given $z_0 \in \mathbb{C}^g$. This shows that $\phi(z_0)$ is defined. By the same method, one generates sufficiently many functions to separate points, and to show injectivity of $d\phi$. See [MAV, pp 29-33] for full details. □

Let us characterize lattices of the form $L = L_\Omega = \mathbb{Z}^g + \Omega\mathbb{Z}^g$, with $\Omega \in \mathbb{H}_g$, in coordinate free language. Let $e_1, \ldots e_g$, be the standard basis of $\mathbb{Z}^g$. We can extend this to basis of L, by taking e_{g+i} to be the ith column of Ω. The vectors $e_1, \ldots, e_{2g}$ form a *real* basis of $\mathbb{C}^g$. Let $E : \mathbb{C}^g \times \mathbb{C}^g \to \mathbb{R}$ be the real bilinear form with matrix

$$\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \tag{4.2}$$

with respect to this basis. By definition, it is skew symmetric $E(v, u) = -E(u, v)$.

Lemma 4.1.3.

(a) $E(u, v) = \operatorname{Im}(u^t(\operatorname{Im}\Omega)^{-1}\bar{v})$

(b) $E(u, v) \in \mathbb{Z}$*, when* $u, v \in L$*.*

(c) $E(iu, iv) = E(u, v)$

(d) $E(iu, v)$ *is symmetric positive definite.*

(e) *There exist a positive definite hermitian form H on $\mathbb{C}^g$, such that $E = \operatorname{Im} H$.*

Proof. Item (a) can be checked by calculation, (b) is clear, and (c) and (d) follow from (a). These conditions show that

$$H(x, y) = E(ix, y) + iE(x, y)$$

satisfies (e). It is worth noting that (c) and (d) also follow from (e). □

Given a lattice $L \subset \mathbb{C}^g$, a nondegenerate skew-symmetric form $E : \mathbb{C}^g \times \mathbb{C}^g \to \mathbb{R}$ satisfying (b), (c) and (d) (or equivalently (b) and (e)) is called a *polarization* or Riemann form. It is called *principal* if in addition $\det E = 1$. This implies, by standard linear algebra arguments, that L possesses a basis such that E is given by (4.2). Since E and H above determine each other, H is also sometimes referred to as the polarization.

Lemma 4.1.4. *If L has a principal polarization, then after choosing suitable bases for $\mathbb{C}^g$ and L, we have $L = L_\Omega$ for some $\Omega \in \mathbb{H}_g$.*

Proof. We will say a bit more about this later on. □

We can thus rephrase theorem 4.1.2 as saying that $\mathbb{C}^g/L$ is an abelian variety if L possesses a principal polarization. In fact, by allowing arbitrary polarizations, we get an if and only if statement.

Theorem 4.1.5 (Riemann, Lefschetz)**.** *$\mathbb{C}^g/L$ is an abelian variety if and only if L possesses a polarization.*

The "if" direction can be proved using theta functions, as above. Let us briefly explain the converse from the viewpoint of complex algebraic geometry [GH], because it explains what E actually means. From algebraic geometry, we know that an embedding $X \subset \mathbb{P}^N$ is determined by the very ample divisor class $H + X \cap$ (hyperplane) or the very ample line bundle $\mathcal{L} = \mathcal{O}_X(1)$. This has the advantage of giving an object on X which doesn't depend on any "external" data. The divisor H determines a homology class $[H] \in H_{2\dim X - 2}(X, \mathbb{Z})$, and by Poincaré duality a cohomology class $[H] \in H^2(X, \mathbb{Z})$. This coincides with the first Chern class $c_1(\mathcal{L}) \in H^2(X, \mathbb{Z})$, which is the basic topological invariant of a line bundle. Since X is a torus, we can identify $H^2(X, \mathbb{Z}) = \wedge^2 Hom(L, \mathbb{Z})$. In other words, $c_1(\mathcal{L})$ can be viewed as an alternating integer valued pairing E on the lattice L. This means that E satisfies condition (b) for a polarization. On the other hand, since $c_1(\mathcal{L})$ is the restriction of $c_1(\mathcal{O}_{\mathbb{P}^N}(1))$, it can be represented by the normalized curvature of the Fubini-Study metric. In particular, it can also be represented by a real differential form, called the Kähler form,

$$\omega = \frac{\sqrt{-1}}{2} \sum_{jk} h_{jk} dz_j \wedge d\bar{z}_k$$

with h_{jk} positive definite hermitian. This can be used to show that E satisfies (e) as well.

From this discussion, we obtain.

Corollary 4.1.6. *Under the identification* $H^2(X,\mathbb{Z}) = \wedge^2 Hom(L,\mathbb{Z})$*, an element is a polarization if and only if is the first Chern class of an ample line bundle.*

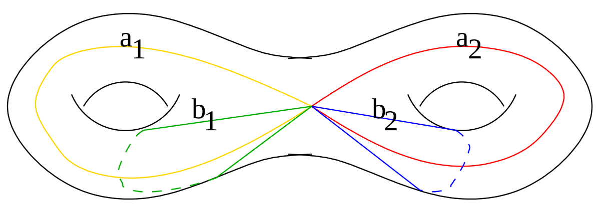
Although polarizations do not traditionally appear in the theory of elliptic curves, they exist and are easy to describe. When X is an elliptic curve $H^2(X,\mathbb{Z}) \cong \mathbb{Z}$. Under this isomorphism, the Chern class of a divisor $c_1(\mathcal{O}(D))$ is just its degree $\deg D$. It is ample, and therefore corresponds to a polarization, if $\deg D$ is positive. And it is principal, when $\deg D = 1$. So X has a unique principal polarization.

4.2 Jacobians

Let X be a nonsingular projective curve of genus g. Recall that

$$g = \dim H^1(X, \mathcal{O}_X) = \dim H^0(X, \Omega^1_X)$$

Earlier, we took the groups on the right to be sheaf cohomology of the sheaves of regular functions/forms on the Zariski topology, we can also (and will) interpret these as the cohomology groups of the sheaves of holomorphic functions/forms on the classical topology. This is justified by Serre's GAGA theorems. We want to explain that g is the same topological genus, which one half the dimension of the first de Rham cohomology group $H^1(X,\mathbb{C})$ of closed C^∞ complex 1-forms modulo exact forms.



Let say that a 1-form α is *harmonic* if in any system of local analytic coordinates $z = x+iy$, $\alpha = df(x,y)$ where f is harmonic in the usual sense, i.e. it lies in the kernel of $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. (This definition is a bit nonstandard, but people familiar with the the usual condition $(d^*d + dd^*)\alpha = 0$ should be able to check the equivalence.) The key fact, that we state without proof, is the Hodge theorem (which is really due to Weyl in the case of Riemann surfaces).

Theorem 4.2.1 (Hodge theorem)**.** $H^1(X,\mathbb{C})$ *is isomorphic to the space of harmonic* 1*-forms.*

Holomorphic 1-forms are harmonic for example, since any such form is locally df with f holomorphic, and basic complex analysis teaches us that holomorphic functions are harmonic. Conversely, a harmonic $(1,0)$-form, i.e. a form locally a multiple of dz, is necessarily holomorphic. We can also see that a $(0,1)$-form (a multiple of $d\bar{z}$) is harmonic if and only if it is a conjugate of a holomorphic 1-form. Thus:

Corollary 4.2.2 (Hodge decomposition)**.** *We have decomposition*

$$H^1(X,\mathbb{C}) = H^{10}(X) \oplus H^{01}(X)$$

where $H^{01}(X) = \overline{H^{10}(X)}$ *and* $H^{10} = H^0(X,\Omega^1_X)$*. In particular,* $\dim H^1(X,\mathbb{C}) = 2g$.

We should explain how to interpret complex conjugation in the above result. To give a conjugation a complex vectors space V is tantamount to finding real vector space $V_{\mathbb{R}}$ and an isomorphism $V_{\mathbb{R}} \otimes \mathbb{C} \cong V$; then $\overline{v \otimes a} = v \otimes \overline{a}$. For $V = H^1(X,\mathbb{C})$, we take $V_{\mathbb{R}} = H^1(X,\mathbb{R})$ to be the de Rham cohomology of real differential forms. Basic facts from topology (the de Rham and universal coefficient theorems), tells us that we can take this further. If $H^1(X,\mathbb{Z}) \cong \mathbb{Z}^{2g}$ is singular cohomology with integer coefficients, then

$$\begin{aligned} H^1(X,\mathbb{C}) &\cong H^1(X,\mathbb{Z}) \otimes \mathbb{C} \\ H^1(X,\mathbb{R}) &\cong H^1(X,\mathbb{Z}) \otimes \mathbb{R} \end{aligned} \tag{4.3}$$

To make this more explicit, note that the dual $Hom(H^1(X,\mathbb{Z}),\mathbb{Z})$ can identified with the homology $H_1(X,\mathbb{Z})$. Elements of this are represented by (sums of) closed smooth loops on X. Given a form α representing a class in $H^1(X,\mathbb{C})$, the map (4.3) sends α to the functional $\gamma \mapsto \int_\gamma \alpha$.

We will also consider the transpose of this map (4.3)

$$H_1(X,\mathbb{Z}) \to H^1(X,\mathbb{C})^*,\ \gamma \mapsto \int_\gamma$$

This restricts to

$$H_1(X,\mathbb{Z}) \to H^0(X,\Omega^1_X)^* \tag{4.4}$$

We define the Jacobian $J(X)$ as quotient

$$J(X) = \frac{H^0(X,\Omega^1_X)^*}{H_1(X,\mathbb{Z})}$$

where we identify $H_1(X,\mathbb{Z})$ with its image.

Proposition 4.2.3. *$J(X)$ is a complex torus.*

Proof. Given $\alpha \in H^1(X,\mathbb{C})$, write $\alpha^{10} \in H^{10}(X)$ and $\alpha^{01} \in H^{01}(X)$ for its components with respect to the Hodge decomposition. Let $p : H^1(X,\mathbb{R}) \to H^{10}(X)$ be defined by $p(\alpha) = \alpha^{10}$. Suppose that $p(\alpha) = 0$. Then $\alpha = \alpha^{10} + \overline{\alpha^{01}} = 0$. Therefore p is injective. It follows that p is an isomorphism, because both space have the same real dimension. Consequently, we can identify the image of (4.4) with the image of $H_1(X,\mathbb{Z})$ in $H^1(X,\mathbb{R})^*$ which is a lattice. □

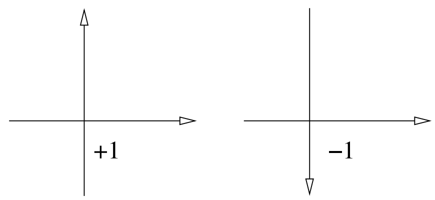
Let $L = H_1(X,\mathbb{Z}) \cong \mathbb{Z}^{2g}$ be the lattice defining $J(X)$. We have an intersection pairing

$$E : L \times L \to \mathbb{Z}$$

where $E(\gamma, \gamma')$ counts the number of times γ intersects γ', with signs. That is if the curves are transverse

$$E(\gamma, \gamma') = \sum_{p \in \gamma \cap \gamma'} \pm 1$$

according to



There are various ways to construct this rigorously. One way is to construct the dual pairing on $H^1(X, \mathbb{Z})$ using the cup product. In terms of the embedding $H^1(X, \mathbb{Z}) \subset H^1(X, \mathbb{R})$, this given by integration

$$E(\alpha, \beta) = \int_X \alpha \wedge \beta$$

Theorem 4.2.4. *E is a principal polarization. Therefore $J(X)$ is an abelian variety.*

Proof. It is already clear from the above formula that E is skew symmetric. Poincaré duality shows that its determinant is $+1$. If we pullback E to $H^{10}(X)$ under the isomorphism p above, we have

$$E(\alpha, \beta) = \int_X (\alpha + \bar{\alpha}) \wedge (\beta + \bar{\beta}) = \int_X \alpha \wedge \bar{\beta} + \bar{\alpha} \wedge \beta$$

It follows that $E(i\alpha, i\beta) = E(\alpha, \beta)$. Finally suppose $\alpha = f(z)dz$ where f is nonzero holomorphic. Since

$$i\alpha \wedge \bar{\alpha} = 2|f(z)|^2 dx \wedge dy$$

we conclude that

$$E(i\alpha, \alpha) > 0$$

□

Finally, let us explain what information $J(X)$ carries. Choose a base point x_0, and define the Abel-Jacobi map

$$AJ : X \to J(X)$$

by

$$AJ(x) = \int_{x_0}^{x} \in H^{10}(X)^* \mod H_1$$

The integral is only defined after choosing a path from x_0 to x, but its image in $J(X)$ does not depend on it. Given a divisor $D = \sum n_i x_i$, we define $AJ(D) = \sum n_i AJ(x_i)$.

Theorem 4.2.5 (Abel-Jacobi)**.** *We have an isomorphism of abelian groups*

$$Cl^0(X) \cong J(X)$$

induced by AJ, where $Cl^0(X)$ is the degree zero part of the divisor class group.

4.3 Siegel modular varieties

If $A_i = V_i/L_i$ are complex tori, we will say that they are *isomorphic* (respectively *isogenous*) if there is a linear isomorphism $\phi : V_1 \to V_2$ such that $\phi(L_1) = L_2$ (resp. $\phi(L_1) \subseteq L_2$).

Lemma 4.3.1. *Isogeny is an equivalence relation.*

Proof. Transitivity and reflexivity are obvious. We only have to prove that isogeny is symmetric. If $\phi : V_1 \to V_2$ is an isomorphism such that $\phi(L_1) \subseteq L_2$, then L_1 is a sublattice of $\phi^{-1}(L_2)$. Therefore $N\phi^{-1}(L_2) \subset L_1$ for some N. This means that $N\phi^{-1}$ is an isogeny in the opposite direction. □

In case these are abelian varieties, an isomorphism in this sense is automatically an isomorphism of algebraic varieties by [GAGA]. If A_i are equipped with polarizations E_i, we say that ϕ is an isomorphism of polarized abelian varieties if ϕ preserves the forms E_i, i.e.

$$E_1(u, v) = E_2(\phi(u), \phi(v))$$

The problem of describing all abelian varieties up to isomorphism does not have a good solution, but the polarized version does. We now describe it.

Lemma 4.3.2. *An abelian variety is isogenous to a principally polarized abelian variety. Any principally polarized abelian variety of dimension g is isomorphic, as a polarized abelian variety, to an abelian variety of the form $A_\Omega = \mathbb{C}^g/L_\Omega$, $L_\Omega = \Omega\mathbb{Z}^g + \mathbb{Z}^g$ with*

$$E = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

for some $\Omega \in \mathbb{H}_g$.

Proof. The first statement is [BL, 4.1.2]. The second is just a restatement of lemma 4.1.4 □

Of course, the Ω in the previous lemma is not unique. Let us introduce the symplectic group. Given a commutative ring R (e.g. $\mathbb{Z}, \mathbb{R}$) we define

$$Sp_{2g}(R) = \left\{ M \in GL_{2g}(R) \mid M^T E M = E \right\}$$

Lemma 4.3.3. *Given* $\Omega \in \mathbb{H}_g$ *and* $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in Sp_{2g}(\mathbb{R})$

$$M \cdot \Omega = (A\Omega + B)(C\Omega + D)^{-1} \in \mathbb{H}_g$$

This defines an action of $Sp_{2g}(\mathbb{R})$ *on* $\mathbb{H}_g$ *which is transitive. The isotropy group of* iI *is*

$$\left\{ \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \mid AB^T = BA^T, AA^T + BB^T = I \right\} \cong U_n(\mathbb{R})$$

where the isomorphism is given by sending

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix} \mapsto A + iB$$

Proof. For M as above, one checks the following identities: A^TC and B^TD are symmetric, and $A^TD - C^TB = I$. After expanding, using the above identities, and canceling, we obtain

$$(C\Omega + D)^T(M \cdot \Omega - (M \cdot \Omega)^T)(C\Omega + D) = \Omega - \Omega^T = 0$$

Therefore $M \cdot \Omega$ is symmetric. Similarly

$$(C\Omega + D)^T(\operatorname{Im} M \cdot \Omega)(C\Omega + D) = \operatorname{Im} \Omega > 0$$

which implies that $\operatorname{Im} M \cdot \Omega$ is positive definite.

Let $\Omega = X + iY \in \mathbb{H}_g$. Since Y is symmetric and positive definite, we can find an $A \in GL_g(\mathbb{R})$ so that $Y = AA^T$. Then $M = \begin{pmatrix} A & X(A^T)^{-1} \\ 0 & (A^T)^{-1} \end{pmatrix}$ sends iI to Ω. The formula for the isotropy group can be checked by calculation. □

Corollary 4.3.4. *Thus* $\mathbb{H}_g \cong Sp_{2g}(\mathbb{R})/U_g(\mathbb{R})$.

Let us now explain the idea for the proof of lemma 4.1.4. Given a principally polarized abelian variety $(V/L, E)$. Choose a symplectic basis $\lambda_1, \ldots, \lambda_{2g}$ for L. A basis is symplectic if E is represented by the matrix (4.2). Use the first g vectors $\lambda_1, \ldots, \lambda_g$ as a basis for V. Then if we write the remaining vectors $\lambda_{g+1}, \ldots, \lambda_{2g}$ in the last basis, we get a $g \times g$ matrix Ω. One can see that the conditions for a polarization force $\Omega \in \mathbb{H}_g$ [BL, §4.2]. So once we fix the initial basis λ_i, Ω is determined. If λ_i' is a different symplectic basis, then we get a different $\Omega' \in \mathbb{H}_g$. The relationship is easy to work out. The change of basis matrix $\lambda_i' = \sum m_{ij}\lambda_j$ is necessarily in $Sp_{2g}(\mathbb{Z})$.

Lemma 4.3.5. $\Omega' = M \cdot \Omega$.

We define

$$A_g = Sp_{2g}(\mathbb{Z})\backslash\mathbb{H}_g = Sp_{2g}(\mathbb{Z})\backslash Sp_{2g}(\mathbb{R})/U_g(\mathbb{R})$$

This is called a *Siegel modular variety*. Although at the moment it is just a set.

Corollary 4.3.6. *There is a natural one to one correspondence between elements of A_g and isomorphism classes of g dimensional principally polarized abelian varieties.*

Next, we study the action of $Sp_{2g}(\mathbb{Z})$ on $\mathbb{H}_g$.

Lemma 4.3.7. *The action of $Sp_{2g}(\mathbb{Z})$ is properly discontinuous. Therefore the quotient is a Hausdorff space.*

Proof. Given compact sets $K_1, K_2 \subset \mathbb{H}_g$, we have to show that $S = \{M \in Sp_{2g}(\mathbb{Z}) \mid M(K_1) \cap K_2 \neq \emptyset\}$ is finite. Let us identify $\mathbb{H}_g = Sp_{2g}(\mathbb{R})/U_g(\mathbb{R})$ as above. Note that the group $U_g(\mathbb{R})$ is compact, so that the projection $p : Sp_{2g}(\mathbb{R}) \to \mathbb{H}_g$ is proper. $M \in Sp_{2g}(\mathbb{Z})$ lies in S if and only if $Mp^{-1}K_1 \cap p^{-1}K_2 \neq \emptyset$ if and only if $M \in T = (p^{-1}(K_1))^{-1}p^{-1}(K_2)$. Now T is compact because it is the image of $K_1 \times K_2$ under $(M_1, M_2) \mapsto M_1^{-1}M_2$. Therefore S is the intersection of a compact set with a discrete set, so it's finite. □

The action has fixed points. The solution, as before, is to pass to a congruence subgroup

$$\Gamma(N) = \ker[Sp_{2g}(\mathbb{Z}) \to Sp_{2g}(\mathbb{Z}/N\mathbb{Z})]$$

Proposition 4.3.8. *If $N \geq 3$, then $\Gamma(N)$ is torsion free.*

Proof. We assume that $\gamma \neq I$ is an element of $\Gamma(N)$ of finite order. We can assume that the order is a prime p, by replacing γ a power. Then by assumption, $I - \gamma = N\phi$ where $\phi \in M_{2g\times 2g}(\mathbb{Z})$. Let ζ be a nontrivial eigenvalue of γ, and let η be the corresponding eigenvalue of ϕ. We have a relation

$$N\eta = 1 - \zeta \tag{4.5}$$

This implies $\eta \in \mathbb{Q}(\zeta)$. Furthermore, η is also an algebraic integer because it satisfies the characteristic polynomial of ϕ. Suppose $p = 2$, then $\zeta = -1$. Equation (4.5) implies $N|2$, which is a contradiction because $N \geq 3$. Now suppose $p \geq 3$. Then ζ is a primitive pth root of unity and, as already noted, η is an algebraic integer in the cyclotomic field $\mathbb{Q}(\zeta)$. Taking the norm of (4.5) with respect to $\mathbb{Q}(\zeta)/\mathbb{Q}$ yields an equality of integers

$$N^{p-1}Norm(\eta) = (1-\zeta)(1-\zeta^2)\ldots(1-\zeta^{p-1}) = p$$

But this is impossible because p is prime. □

A consequence of the proposition is that $\Gamma(N)$, with $N \geq 3$, acts freely on $\mathbb{H}_g$. So the quotient

$$A_{g,N} = \Gamma(N)\backslash\mathbb{H}_g$$

can be seen to be a manifold. In more detail, define $\mathcal{O}_{\mathcal{A}_g}(U)$ (and $\mathcal{O}_{\mathcal{A}_{g,N}}(U)$) as to correspond to invariant holomorphic functions on the preimage $\tilde{U} \subset \mathbb{H}_n$. Since the action of $\Gamma(N)$ is free, we find that

Proposition 4.3.9. *When $N \geq 3$, $A_{g,N}$ is a complex manifold.*

A_g is a quotient of $A_{g,N}$ by the finite group $Sp_{2g}(\mathbb{Z}/N\mathbb{Z})$. Therefore

Corollary 4.3.10. *$(A_g, \mathcal{O}_{A_g})$ is a normal analytic space.*

4.4 Siegel modular varieties are moduli spaces

As we did for elliptic curves, we want to upgrade corollary 4.3.6 to a more precise statement. Let us formulate it more generally for $A_{g,N}$. Given $A_\Omega = \mathbb{C}^g/\Omega\mathbb{Z}^g + \mathbb{Z}^g$, the standard basis of the lattice mod N is called a level N-structure. One can see that $M \in \Gamma(N)$ preserves this basis. In general, a level N-structure on a principally polarized abelian variety A is a basis of $H_1(A, \mathbb{Z}/N\mathbb{Z})$ which is symplectic with respect to the form induced by the polarization. In more algebraic terms, it can also be taken as basis of the N-torsion, which is symplectic in the appropriate sense.

Theorem 4.4.1. *$A_{g,N}$ is the coarse moduli space for principally polarized g dimensional abelian varieties with N-structure. When $N \geq 3$, this is a fine moduli space.*

The last statement means that there is a universal family of abelian varieties with the above structure. We will now outline the construction. Given $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in Sp_{2g}(\mathbb{R})$ and $(\Omega, z) \in \mathbb{H}_g \times \mathbb{C}^g$, define

$$M \cdot (\Omega, z) = (M \cdot \Omega, ((C\Omega + D)^T)^{-1} z)$$

Lemma 4.4.2. *This defines an action.*

Proof. [MT, p 177]. □

We define a real linear isomorphism

$$i_\Omega : \mathbb{R}^{2g} \to \mathbb{C}^g$$

given by sending $(v_1, v_2) \in \mathbb{R}^g \times \mathbb{R}^g$ to $\Omega v_1 + v_2$. If $\lambda \in \mathbb{Z}^{2g}$, let

$$\lambda \cdot (\Omega, z) = (\Omega, z + i_\Omega(\lambda))$$

Let $\tilde{\Gamma}$ (resp. $\tilde{\Gamma}(N)$) be the subgroup of the group of holomorphic automorphisms of $\mathbb{H}_g \times \mathbb{C}^g$ generated by $Sp_{2g}(\mathbb{Z})$ (resp. $\Gamma(N)$) and $\mathbb{Z}^{2g}$. A calculation shows that $\mathbb{Z}^{2g}$ is a normal subgroup of $\tilde{\Gamma}$. It follows that $\tilde{\Gamma}$ is a so called semidirect product $Sp_{2g}(\mathbb{Z}) \ltimes \mathbb{Z}^{2g}$. This means that we have a split exact sequence

$$1 \to \mathbb{Z}^{2g} \to \tilde{\Gamma} \to Sp_{2g}(\mathbb{Z}) \to 1$$

In more explicit terms, $\tilde{\Gamma}$ is isomorphic to the cartesian product $Sp_{2g}(\mathbb{Z}) \times \mathbb{Z}^{2g}$, with multiplication

$$(g_1, a_1)(g_2, a_2) = (g_1 g_2, a_1 + g_1 a_2)$$

With the help of this structure, we can see that the action of this group on $\mathbb{H}_g \times \mathbb{C}^g$ is properly continuous, and free when restricted to $\tilde{\Gamma}(N)$, for $N \geq 3$. Consequently the quotient

$$U_{g,N} = \tilde{\Gamma}(N) \backslash \mathbb{H}_g \times \mathbb{R}^{2g}$$

is a complex manifold. Projection on the first factor yields a holomorphic map $\pi : U \to A_{g,N}$. The fibre over a point corresponding to Ω is the abelian variety A_Ω. This is our desired universal family.

An application of Baily-Borel shows that $A_{g,N}$ is a quasiprojective variety. Using a completely different construction, Mumford [GIT] proved that

Theorem 4.4.3 (Mumford)**.** *$A_{g,N}$ is the set of complex points of a quasiprojective scheme over* $\operatorname{Spec}\mathbb{Z}$*. This is a coarse moduli space for all N. It is fine, and smooth over* $\operatorname{Spec}\mathbb{Z}[1/N, \exp(2\pi i/N)]$*, when $N \geq 3$.*

The fact that this space exists over $\mathbb{Z}$ has a number of important arithmetical applications. For example, Faltings [F1] used this in an essential way in his original proof of the Mordell conjecture:

Theorem 4.4.4 (Faltings)**.** *A smooth projective curve of genus ≥ 2 defined over a number field K, has only finitely many K-rational points.*

Chapter 5

The endomorphism algebra

5.1 Endomorphisms of elliptic curves

A *homomorphism* between abelian varieties $f : V/L \to W/M$ is given by a $\mathbb{C}$-linear map $F : V \to W$ such that $F(L) \subseteq M$. This is called an endomorphism if the abelian varieties are the same. Our goal is to study the ring of endomorphisms $End(A)$, of an abelian variety. We start with elliptic curves.

Theorem 5.1.1. *Let $E = \mathbb{C}/\mathbb{Z} + \mathbb{Z}\tau$, then either*

1. *$End(E) = \mathbb{Z}$ or*
2. *$\mathbb{Q}(\tau)$ is an imaginary quadratic field, and $End(E)$ is an order in $\mathbb{Q}(\tau)$ i.e. a finitely generated subring such that $End(E) \otimes \mathbb{Q} = \mathbb{Q}(\tau)$.*

Proof. Let $L = \mathbb{Z} + \mathbb{Z}\tau$. Then $End(E)$ can be identified with $R = \{\alpha \in \mathbb{C} \mid \alpha L \subseteq L\}$. For $\alpha \in R$, there are integers ab, c, d such that

$$\alpha = a + b\tau,\ \alpha\tau = c + d\tau$$

By Cayley-Hamilton, or direct calculation, we see that

$$\alpha^2 - (a + d)\alpha + ad - bc = 0$$

Therefore R is an integral extension of $\mathbb{Z}$.

Suppose that $R \neq \mathbb{Z}$, and choose $\alpha \in R$ but $\alpha \notin \mathbb{Z}$. Then eliminating α from the top two equations yields

$$b\tau^2 - (a - d)\tau - c = 0$$

Therefore $\mathbb{Q}(\tau)$ is quadratic, and necessarily imaginary quadratic because τ is not real. Furthermore, $R \subset \mathbb{Q}(\tau)$ is an order. □

5.2 Poincaré reducibility

A homomorphism between abelian varieties $f : V/L \to W/M$ f is called an *isogeny* if F is an *isomorphism*, and an isomorphism if in addition $F(L) = M$. Isomorphisms are always bijections, while isogenies a finite to one surjections. For example, multiplication by a nonzero integer $n : V \to V$ induces an isogeny, which is not an isomorphism unless $n = \pm 1$. Two abelian varieties X and Y are called isogenous if there exists an isogeny from X to Y. Recall

Lemma 5.2.1. *Isogeny is an equivalence relation.*

We gave a direct proof earlier. We can give another proof, by interpreting isogeny in a fancier way. The collection of abelian varieties and homomorphisms forms an additive category $AbVar$. We can form a new category $AbVar_{\mathbb{Q}}$ with the same objects but with morphisms given by $Hom_{\mathbb{Q}}(X, Y) = Hom(X, Y) \otimes \mathbb{Q}$. We also set $End(X) = Hom(X, X)$ and $End_{\mathbb{Q}}(X) = End(X) \otimes \mathbb{Q}$. These are both not necessarily commutative rings. The previous lemma is now an immediate consequence of the observation:

Lemma 5.2.2. *Two abelian varieties are isogenous if and only if they are isomorphic in $AbVar_{\mathbb{Q}}$.*

Corollary 5.2.3. *$End_{\mathbb{Q}}(X)$ depends only on the isogeny class of X.*

Theorem 5.2.4 (Poincaré)**.** *If $X \subset Y$ is an injective homomorphism of abelian varieties, then Y is isogenous to a product X with another abelian variety.*

Proof. Suppose that $Y = V/L$ then $X = W/L \cap W$ for some subspace $W \subset V$ Let $W^{\perp}$ be the orthogonal complement with respect to a polarization H. Then this is also the orthogonal complement with respect to $E = \operatorname{Im} H$ by lemma 4.1.3. Equivalently, $W^{\perp}$ is the kernel of the map $v \mapsto E(v, -)$. Since this transformation can be represented by a rational matrix with respect to a basis of L, this implies that $\dim_{\mathbb{Q}} L_{\mathbb{Q}} \cap W^{\perp} = \dim_{\mathbb{R}} W^{\perp}$. Therefore $L \cap W^{\perp}$ is a lattice in $W^{\perp}$, so we can form torus $Z = W^{\perp}/L \cap W^{\perp}$. This is an abelian variety polarized by the restriction of H. The identity map $W \oplus W^{\perp} = V$ defines an isogeny $X \times Z \to Y$. □

Remark 5.2.5. *When Y has a polarization which restricts to a principal polarization of X, the map $X \times Z \to Y$ can be seen to be an isomorphism, cf. [BL, 5.3.13]. This is not true in general.*

An abelian variety is *simple* if it contains no nontrivial abelian subvarieties.

Corollary 5.2.6. *An abelian variety is isogenous to a product of simple abelian varieties.*

We turn now to the structure of the endomorphism ring $End_{\mathbb{Q}}(X)$. In general, it is noncommutative. The following is a standard argument in representation theory.

Theorem 5.2.7. *If X is simple, then $End_{\mathbb{Q}}(X)$ is a finite dimensional division algebra over $\mathbb{Q}$. In general, $End_{\mathbb{Q}}(X)$ is a product of matrix algebras over finite division algebras over $\mathbb{Q}$.*

Proof. The finite dimensionality is clear from construction, since $End_{\mathbb{Q}}(X) \subset End(L \otimes \mathbb{Q})$ where L is the lattice. Suppose that X is simple and that $f \in End_{\mathbb{Q}}(X)$ is nonzero. We have to show that f has an inverse. After replacing f by nf, we can assume that it is a homomorphism $f : X \to X$. It is enough to show that it is an isogeny. Since $f(X) \subset X$ is nonzero abelian subvariety, it follows that $f(X) = X$. Consider $\ker(f) \subset X$. It must be finite, since otherwise the connected component of the identity would give a nonzero abelian subvariety. It follows that f is an isogeny.

For the second statement, there is no loss in assuming that $X = \prod X'_i$ where X'_i simple. We can arrange this as $X = \prod X_i^{n_i}$ where X_i and X_j are nonisogenous when $i \neq j$. Given $\phi : X \to X$, we can decompose it as a product of morphisms $\phi_{ij} : X'_i \to X'_j$. Since both X'_i and X'_j are simple, ϕ_{ij} is either 0 or an isogeny. Thus we can decompose ϕ as product of matrices with values in $D_i = End_{\mathbb{Q}}(X_i)$. In other words, we have an inclusion

$$End_{\mathbb{Q}}(X) \hookrightarrow \prod Hom_{\mathbb{Q}}(X'_i, X'_j) = \prod Mat_{n_i \times n_i}(D_i)$$

This is clearly surjective as as well. ∎

An algebra of the above type is called semisimple.

5.3 The Rosati involution

There is an extra bit of structure which will be play a very important role. Given an algebra R over a field. An involution is a map $r \mapsto r^*$ which is linear over the field, such that $(rs)^* = s^* r^*$. For example, transpose gives an involution of on the algebra of matrices.

Let $X = V/L$ be an abelian variety with polarization H. The adjoint with respect to H:

$$H(Ax, y) = H(x, A^* y)$$

defines an involution on $End(V)$. The algebra $End_{\mathbb{Q}}(X)$ sits naturally inside this. It can be identified with the endomorphisms which preserve the rational lattice $L_{\mathbb{Q}} = L \otimes \mathbb{Q}$.

Theorem 5.3.1. *The subring $End_{\mathbb{Q}}(X) \subset End(V)$ is stable under the involution $*$.*

Proof. If $A \in End(L_{\mathbb{Q}})$ define $A^\dagger \in End(L_{\mathbb{Q}})$ to be the adjoint with respect to $E = \operatorname{Im} H$ i.e. $E(Ax, y) = E(x, A^\dagger y)$. This is defined because E is nonsingular. Given $A \in End_{\mathbb{Q}}(X)$, it preserves $L_{\mathbb{Q}}$, so we can form $A^\dagger \in End(L_{\mathbb{Q}})$. This coincides with the usual adjoint $A^* \in End(V)$ because $\operatorname{Im} H(Ax, y) = \operatorname{Im} H(x, A^* y)$. Therefore A^* preserves the rational lattice $L_{\mathbb{Q}}$, and thus defines an element of $End_{\mathbb{Q}}(X)$. ∎

The restriction of $*$ to $End_{\mathbb{Q}}(X)$ is called the *Rosati involution*. Although the construction would seem to be based on a linear algebra trick, there is a way to make it more geometric. Let V^* be the space of complex *antilinear* maps $V \to \mathbb{C}$. This means that $f(av_1 + a_2v_2) = \bar{a}_1 f(v_1) + \bar{a}_2 f(v_2)$. This is can be understood as complex conjugate of the usual dual. Let $L^* \subset V^*$ denote the subset of those maps which are integer valued on L. The quotient $\hat{X} = V^*/L^*$ is a torus, which is in fact an abelian variety. The map $v \mapsto H(v, -)$ induces an isogeny ϕ_H between X and its dual $\hat{X} = V^*/L^*$. Thus we have an isomorphism $\Phi : End_{\mathbb{Q}}(X) \cong End_{\mathbb{Q}}(\hat{X})$. An endomorphism $A : X \to X$ induces a dual endomorphism $\hat{A} : \hat{X} \to \hat{X}$. Then $A^* \in End_{\mathbb{Q}}(X)$ is $\Phi^{-1}(\hat{A})$. All of this can be described geometrically as follows:

Theorem 5.3.2. *There is an isomorphism $Pic^0(X) \cong \hat{X}$ under which $\hat{A}$ corresponds to the homomorphism $Pic^0(X) \to Pic^0(X)$ given by $L \mapsto A^*L$. If M is an ample line bundle on X with the H the first Chern class (as discussed in section 4.1), then there is an isogeny $\phi_H : X \to \hat{X}$ is given by $\phi_H(x) = T_x^* M \otimes M^{-1} \in \hat{X}$. The Rosati involution is given by*

$$A^* = \phi_H^{-1} \hat{A} \phi_H$$

Proof. [BL, MAV]. □

Given any finite dimensional $\mathbb{Q}$-algebra R, and element r defines a vector space endomorphism of R by left multiplication. This is the so called regular representation. Thus we have a well defined trace $Tr(r) \in \mathbb{Q}$. An involution $*$ on R is called *positive* if $Tr(r^*r) > 0$ when $r \neq 0$. Transpose on the algebra of matrices has this property.

Theorem 5.3.3. *The Rosati involution is positive.*

Proof. Let D be an ample divisor representing the polarization. One has that

$$Tr(r^*r) = \frac{2g}{D^g}(D^{g-1} \cdot r^*D) > 0$$

See [BL, p 117]. □

5.4 Division rings with involution

In the first section, we showed that $End_{\mathbb{Q}}$ of an elliptic curve was either $\mathbb{Q}$ or an imaginary quadratic field. In higher dimensions, things are more complicated, but that they can be understood. Given a simple abelian variety X, $End_{\mathbb{Q}}(X)$ is a finite dimensional division algebra with a positive involution. Our goal is to describe all such rings with involution. Over $\mathbb{R}$, things are much are easier. There are only two (finite dimensional) division algebras over it, the complex numbers $\mathbb{C}$ and the quaternions $\mathbb{H} = \mathbb{R} \oplus \mathbb{R}i \oplus \mathbb{R}j \oplus \mathbb{R}k$ with $i^2 = j^2 = -1$ and $ij = -ji = k$. Both of these algebras have a positive involution given by ordinary complex conjugation and quaternionic conjugation $(x+yi+zj+wk)^* =$

$x-yi-zj-wk$. The construction of quaternions can be generalized to an algebra given by replacing $\mathbb{R}$ by an arbitrary field F, and by modifying the relations to $i^2 = a$, $j^2 = b$ and $ij = -ji = k$ for $a, b \in F^*$. This algebra is usually denoted by the Hilbert symbol $\left(\frac{a,b}{F}\right)$. This carries an involution defined as above. There are two possibilities, either $\left(\frac{a,b}{F}\right)$ is a division algebra, or it is the algebra of 2×2 matrices, in which case, we say that it splits. The algebra splits precisely when the quadratic form $ax^2 + by^2 = 1$ has a solution over F. Therefore, when $F = \mathbb{R}$, $\left(\frac{a,b}{F}\right)$ is nonsplit, and thus $\mathbb{H}$, if and only if $a, b < 0$.

Over $\mathbb{Q}$, there is a classification of division algebras with positive involution.

Theorem 5.4.1 (Albert)**.** *The set of finite dimensional division algebras $(D, *)$ over $\mathbb{Q}$ with a positive involution are exactly the ones described below (written out of the traditional order).*

Type I. *$D = F$ is a totally real number field F (this means that all complex embeddings lie in $\mathbb{R}$) We give this the trivial involution $x^* = x$.*

Type III. *D is a quaternion algebra $\left(\frac{a,b}{F}\right)$ over a totally real number field F with $a, b \in F$ totally negative. The involution is the standard one.*

Type II. *D is a quaternion algebra over a totally real number field F which splits when extended to $\mathbb{R}$ under any embedding $F \hookrightarrow \mathbb{R}$ The involution becomes conjugate to the transpose on the matrix algebra under each isomorphism $D \otimes_F \mathbb{R} \cong M_2(\mathbb{R})$.*

Type IV. *D is a division algebra whose centre is a CM field F, i.e. a quadratic extension $F = K(\sqrt{-D})$ of a totally real field K with D totally positive. The involution restricts to $x + y\sqrt{-D} \mapsto x - y\sqrt{-D}$ on F.*

Proof. We refer to [MAV, pp 193-202] for the proof, and for a more detailed description in case IV. □

Quaternion algebras as in II (resp. III) are also said to be totally indefinite (resp. definite).

Corollary 5.4.2. *The endomorphism algebra of a simple abelian variety must be one of the above 4 types; the abelian variety is labelled accordingly.*

The statement in the corollary can be sharpened somewhat.

Theorem 5.4.3. *Suppose that A is a simple abelian variety of dimension g. Let $D = End_{\mathbb{Q}}(A)$, K the centre, and $e = [K : \mathbb{Q}]$. Then*

1. *$e|g$ (type I)*
2. *$2e|g$ (types II and III)*

3. *$e_0 d | g$ (type IV, where $e_0 = [F : \mathbb{Q}]$ and $d^2 = \dim_{\mathbb{Q}} D$ – it's always a square).*

Proof. We refer to [BL, §5.5] for the details of the proof. We will be content to prove a weaker result that $\dim_{\mathbb{Q}} D | 2g$. To see this, observe that the $End(X)$-action makes the lattice $H_1(A, \mathbb{Q})$ into a D-module, which is necessary free of rank of rank r say, because D is a division algebra. It follows that $r \dim_{\mathbb{Q}} D = \dim_Q H_1(A, \mathbb{Q})$. □

We will see later that all of the categories I-IV occur for abelian varieties, and almost all of the subcases. The idea is easy to explain for elliptic curves.. In theorem 5.1.1, we saw that an elliptic curve $E = \mathbb{C}/\mathbb{Z}+\mathbb{Z}\tau$ has either $End_{\mathbb{Q}}(E) = \mathbb{Q}$ (special case of type I) or $End_{\mathbb{Q}}(E)$ imaginary quadratic (special case of type IV). Furthermore, in the second case, $End_{\mathbb{Q}}(E) = \mathbb{Q}(\tau)$. The converse is simple.

Lemma 5.4.4. *Given $\mathbb{Q}$ or an imaginary quadratic field, it arises as above.*

Proof. To build an elliptic curve with E with $End_{\mathbb{Q}}(E) = \mathbb{Q}(\sqrt{-d})$ we can use $E = \mathbb{C}/\mathbb{Z} + \mathbb{Z}\sqrt{-d}$. For $End_{\mathbb{Q}}(E) = \mathbb{Q}$, suffices to take $\mathbb{C}/\mathbb{Z} + \mathbb{Z}\tau$ with $\mathbb{Q}(\tau)$ not imaginary quadratic. For example, we can take τ transcendental. □

It is clear that "most" E have $End_{\mathbb{Q}}(E) = \mathbb{Q}$. Making this idea work in higher dimensions will require some understanding of moduli spaces.

Chapter 6

Introduction to Shimura varieties

6.1 Hilbert modular varieties.

A g-dimensional abelian variety A is said to have *real multiplication* if it is of type I. This means that $K = End_{\mathbb{Q}}(A)$ is a totally real field with degree $[K : \mathbb{Q}]$ dividing g. So the maximum possible degree is g. The endomorphism algebra $R = End(A)$ is an *order* in K. This means that R is a subring, which is also an $\mathcal{O}_K$-lattice.

We would like describe all abelian varieties with real multiplication by a field with maximal degree. Fix a totally real field K, of degree g. We will, in fact, describe all principally polarized g dimensional abelian varieties A with $End(A) \supseteq \mathcal{O}_K$. As a $\mathbb{Z}$-module, $\mathcal{O}_K \cong \mathbb{Z}^g$. For each vector $\tau = (\tau_j) \in \mathbb{H}^g$ define $L_\tau \subset \mathbb{C}^g$ to be the image of $\mathcal{O}_K^2$ under the map

$$\iota_\tau(\alpha, \beta) = (\sigma_j(\alpha)\tau_j + \sigma_j(\beta))$$

where $\sigma_1, \ldots, \sigma_g : K \to \mathbb{R}$ are the different embeddings.

Proposition 6.1.1. *L_τ is a lattice, and the quotient $A_\tau = \mathbb{C}^g/L_\tau$ is an abelian variety with $\mathcal{O}_K \subseteq End(A_\tau)$.*

Proof. We note that $K \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{R}^g$ where the projections to the factors are the σ_j. It follows that $\mathcal{O}_K \subset K \subset \mathbb{R}^g$ is lattice, and therefore so is $L_\tau \subset \mathbb{C}^g$. Thus A_τ is a torus. Consider the Hermitian form

$$H(u, v) = \sum \frac{u_j \bar{v}_j}{\operatorname{Im} \tau_j}$$

We claim that this is a polarization. It is clearly positive definite. It remains to show that the imaginary part $\operatorname{Im} H$ is integer valued on the lattice. Let $E_{std} : \mathcal{O}^2 \times \mathcal{O}^2 \to \mathbb{Z}$ be the pairing defined by

$$E_{std}(\alpha_1, \alpha_2; \beta_1, \beta_2) = \operatorname{tr}(\alpha_1\beta_2 - \alpha_2\beta_1) \tag{6.1}$$

One can check that if $(\alpha_1, \alpha_2), (\beta_1, \beta_2) \in \mathcal{O}_K^2$, then

$$\operatorname{Im} H(\iota_\tau(\alpha_1, \alpha_2), \iota_\tau(\beta_1, \beta_2)) = E_{std}(\alpha_1, \alpha_2; \beta_1, \beta_2)$$

So it is a polarization as claimed. Therefore A_τ is an abelian variety. Furthermore, one can check that the determinant of E_{std} is 1, so this is a principal polarization.

Finally, we have an embedding $\mathcal{O}_K \subset M_{g\times g}(\mathbb{C})$ which sends α to the diagonal matrix with entries $\sigma_j(\alpha)$. L_τ is stable under the resulting $\mathcal{O}_K$-action. Therefore $\mathcal{O}_K \subset End(A_\tau)$. □

The Hilbert modular group is $\Gamma_K = SL_2(\mathcal{O}_K)$. We can embed this into $SL_2(\mathbb{R})^g$ by $M \mapsto (\sigma_j(M))$. Thus we get an action of Γ_K on $\mathbb{H}^g$. The quotient $\Gamma_K \backslash \mathbb{H}^g$ is called a Hilbert modular variety. This is a moduli space:

Theorem 6.1.2. *The points of $\Gamma_K \backslash \mathbb{H}^g$ are in one to one correspondence with isomorphism classes of the following sets of data:*

1. *A g-dimensional polarized abelian variety (A, E)*
2. *An inclusion $\mathcal{O}_K \subseteq End(A)$*
3. *An $\mathcal{O}_K$-module isomorphism $H_1(A, \mathbb{Z}) \cong \mathcal{O}_K^2$ taking E to E_{std}.*

In one direction, we note that by construction, A_τ carries an inclusion $\mathcal{O}_K \subseteq End(A_\tau)$ and an isomorphism

$$H_1(A_\tau, \mathbb{Z}) \cong L_\tau \cong \mathcal{O}_K^2$$

such that E_{std} polarizes A_τ.

Lemma 6.1.3. *If $\tau, \tau' \in \mathbb{H}^g$ lie in the same Γ_K-orbit, then then there is an isomorphism $A_\tau \cong A_{\tau'}$ compatible with the identifications $H_1(A_?, \mathbb{Z}) \cong \mathcal{O}_K^2$.*

Proof. Suppose that $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\tau' = M \cdot \tau$. Multiplication by the diagonal matrix D with entries $\sigma_j(c)\tau_j + \sigma_j(d)$ gives an isomorphism $\mathbb{C}^g \cong \mathbb{C}^g$ taking $L_{\tau'}$ to L_τ. The last part follows from the diagram

$$\begin{array}{ccc} \mathbb{C}^g/DL_{\tau'} & \xrightarrow{1} & \mathbb{C}^g/L_\tau \\ \uparrow & & \uparrow \\ \mathcal{O}_K^2 & \xrightarrow{M} & \mathcal{O}_K^2 \end{array}$$

□

6.2 Shimura varieties of PEL type

Shimura [S1] gave a more general construction of moduli spaces of abelian varieties with extra structure, which includes Hilbert modular varieties as special cases. The input consists of a semisimple $\mathbb{Q}$-algebra D (not necessarily a division algebra) with positive involution $*$, centre K and $*$-fixed subfield F. In addition, we have V a left D-module with a nondegenerate alternating $\mathbb{Q}$-bilinear form $E : V \times V \to \mathbb{Q}$ such that

$$E(bu, v) = E(u, b^*v),\ b \in D, u, v, \in V$$

We refer to (D, V, E) as a PE (Polarization-Endomorphism) datum. In the literature, one often adds a lattice in V, and a suitable Level structure, hence one speaks of PEL data. We can now formulate the basic moduli problem:

Given such data, parameterize all polarized abelian varieties (X, E) with

1. *an inclusion $End_{\mathbb{Q}}(X) \supseteq D$ of algebras such that the Rossati involution agrees with the given involution,*
2. *and an isomorphism $H_1(X, \mathbb{Q}) \cong V$ compatible with E and the D-module structure.*

Shimura [S1] proceeds to construct the moduli spaces on a case by case basis. We consider only the first case in detail. We want to parameterize abelian varieties of type I. The PE datum consists of $D = K$ a totally real field of degree d over $\mathbb{Q}$, with trivial involution, and $V = K^{2m}$ with standard symplectic form $E_{std} \oplus E_{std} \oplus \dots$ (6.1). We want to parameterize principally polarized abelian varieties of dimension $g = md$ satisfying the above conditions. This is a generalization of Hilbert modular varieties, which correspond to the case $m = 1$. Let $H = (\mathbb{H}_m)^d$. We have d distinct embeddings $\sigma_i :\ D \to \mathbb{R}$. The group $\Gamma = Sp_{2m}(\mathcal{O}_K)$ acts on H through the homomorphism $\sigma : \Gamma \to (Sp_{2m}(\mathbb{R}))^d$ induced by (σ_i). We claim that the quotient $\Gamma\backslash H$ parameterizes the abelian varieties of the above type. To each m-tuple $\Omega = (\Omega_i) \in H$, we define the subgroup $L_\Omega \subset \bigoplus_i^d \mathbb{C}^m \cong \mathbb{C}^g$ as the image of $\mathcal{O}_D^m \times \mathcal{O}_D^m$ under the map

$$(\alpha, \beta) \mapsto (\Omega_i \sigma_i(\alpha) + \sigma_i(\beta))_{i=1,\dots,d}$$

where $\sigma_i : K \to \mathbb{R}$ are the various embeddings. One sees that this is a lattice, and the quotient $\mathbb{C}^n/L_\Omega$ is an abelian variety satisfying the desired conditions. One application of this construction is the following.

Theorem 6.2.1 (Shimura)**.** *The set of $\Omega \in H$ for which $End_{\mathbb{Q}}(A_\Omega) \neq K$ is a countable union of proper analytic subsets.*

From the Baire category theorem, we deduce

Corollary 6.2.2. *Every totally real field is the endomorphism algebra of some abelian variety.*

Next, we look at a couple of additional examples, which are very special cases of type II and IV. First, let us suppose that $K = \mathbb{Q}$. Fix a totally indefinite quaternion division algebra D over $\mathbb{Q}$. Recall that this means that $D \otimes_{\mathbb{Q}} \mathbb{R} = Mat_{2\times 2}(\mathbb{R})$. For example, $D = \left(\frac{-1,p}{F}\right)$ is such an algebra when p is a prime such that -1 is not a square mod p. For (V, E), we take $V = D$ with $E(x, y) = \mathrm{tr}_{D/\mathbb{Q}}(x\alpha y)$ for some $\alpha \in D^*$ with $\alpha^* = -\alpha$. Choose a lattice $L \subset D$, L can be viewed as a lattice in $Mat_{2\times 2}(\mathbb{R})$ via the embedding $D \subset Mat_{2\times 2}(\mathbb{R})$. Given $\tau \in \mathbb{H}$, L generates a lattice $L \begin{pmatrix} \tau \\ 1 \end{pmatrix} \subset \mathbb{C}^2$. Let

$$X_\tau = \mathbb{C}^2 / L \begin{pmatrix} \tau \\ 1 \end{pmatrix}$$

This is an abelian variety with $End_{\mathbb{Q}}(X_\tau) \supseteq D$. If $\Gamma \subset D^*$ is the subgroup stabilizing the lattice L, the isomorphism class of X_τ depends on the orbit of τ in $\Gamma\backslash\mathbb{H}$. Unlike the case of modular curves, the so called Shimura curve $\Gamma\backslash\mathbb{H}$ is already compact [S2, p 244].

Let us reconsider Picard modular surfaces, but from this viewpoint [LR]. Let $D = F$ be a imaginary quadratic field with discriminant Δ. Note that $\sqrt{\Delta} \in D$. The involution is conjugation, so $F_0 = \mathbb{Q}$. The PE datum consists of $V = F^3$ and E the imaginary part of a hermitian form H with matrix

$$\begin{pmatrix} 0 & 0 & \frac{1}{\sqrt{\Delta}} \\ 0 & 1 & 0 \\ \frac{-1}{\sqrt{\Delta}} & 0 & 1 \end{pmatrix}$$

Let G be the special unitary group of the form H viewed as an algebraic group over $\mathbb{Q}$. Over $\mathbb{R}$, H as signature $(2, 1)$, so $G(\mathbb{R}) = SU(2, 1)$. This acts on the complex 2-ball B. The Picard modular surface is the quotient $\Gamma\backslash B$, where $\Gamma \subset G$ is the subgroup stabilizing the lattice $\mathcal{O}_D^3 \subset V$.

6.3 Representation theoretic viewpoint

Let us continue the discussion of PE moduli problems, but from a more representation theoretic view point, basically following a combination of Mumford [M] and Deligne [D1].

Fix a PE datum (D, V, E). We can regard the D-module V as a $\mathbb{Q}$-vector space through the inclusion $\mathbb{Q} \subset D$. Since it carries nondegenerate symplectic pairing, its dimension is necessarily even, call it $2g$. The group

$$G = \{g \in GL_K(V) \mid g \text{ is } D\text{-linear, } E(gx, gy) = E(x, y)\} \tag{6.2}$$

is an algebraic group over $\mathbb{Q}$. This means that it is subgroup of $GL(V)$ which is also Zariski closed. For any field K, the set of K-rational points

$$G(K) = \{g \in GL_K(V \otimes_{\mathbb{Q}} K) \mid g \text{ is } D \otimes K\text{-linear, } E(gx, gy) = E(x, y)\}$$

also forms a group. When $K = \mathbb{R}$, $V_\mathbb{R} = V \otimes_\mathbb{Q} \mathbb{R}$ is a real vector space, and $G(\mathbb{R}) \subset GL(V_\mathbb{R})$ is Zariski closed, so we can view this as a Lie group. Let $G(\mathbb{R})^o$ denote the connected component of the identity with respect to the classical topology. This embeds into $Sp_{2g}(\mathbb{R})$ as a subgroup.

Choose an element $J_0 \in G(\mathbb{R})$ such that $J_0^2 = -I$. This turns $V_\mathbb{R}$ into a complex vector space V_0, where $a + bi$ acts by $aI + bJ_0$. Let $U(1) \subset \mathbb{C}^*$ denote the unit circle. Then we can see that $h(a + bi) = aI + bJ$ defines a homomorphism $h : U(1) \to G(\mathbb{R})$ of real algebraic groups such that $h(i) = J_0$. Since $J_0 \in G(\mathbb{R})$, we get $E(ix, iy) = E(x, y)$ which is one of the conditions for being a polarization. However, *we have to impose* the remaining condition that

(*) $E(ix, y) = E(J_0 x, y)$ is positive definite.

Then we can see that for any lattice $L \subset V \subset V_\mathbb{R}$ such that E takes integer values on L, $A_0 = V_0/L$ becomes an abelian variety polarized by E such that $D \subseteq End_Q(A_0)$. This is exactly what we wanted, however it is just on example. To generate a family of examples, let $J_g = g^{-1} J_0 g$ for $g \in G(\mathbb{R})$. The following is immediate

Lemma 6.3.1. *$J_g^2 = -I$ and $E(J_g x, y)$ is positive definite.*

Therefore we get another abelian variety A_g, constructed as above using J_g. We can see that this family is parameterized by $X = G(\mathbb{R})^o/K$, where

$$K = \{g \in G(\mathbb{R})^o \mid g^{-1} J_0 g = J_0\} = \{g \in G(\mathbb{R}) \mid J_0 g J_0^{-1} = g\}$$

Lemma 6.3.2. *K is a compact subgroup.*

Proof. One checks that K lies in the unitary group preserving the hermitian form $H(x, y) = E(J_0 x, y) + iE(x, , y)$. The unitary group is compact. □

It follows that K is the group fixed by the involution $Ad(g) := J_0 g J_0^{-1}$. An involution with the property that the fixed group is compact is called a *Cartan involution*. It can be shown that K is a maximal compact subgroup.

Now we have a family of abelian varieties parameterized by the manifold X. Ultimately, we would like both the parameter space and the family to be algebraic varieties. Let us show how to make the space X into a complex manifold. We just the briefest indication of how this is done, referring to [Mi] for further details. Let $\mathcal{G}$ denote the Lie algebra of $G(\mathbb{R})$. This is the tangent space at the identity. $G(\mathbb{R})$ acts on itself by conjugation, and the derivative of this gives he adjoint representation $ad : G(\mathbb{R}) \to GL(\mathcal{G})$. In particular, $ad(J_0) = ad(h(i))$ will act on $\mathcal{G}$. Let $\mathcal{K}$ and $\mathcal{P}$ denote the $+1$ and -1 eigenspaces of $ad(J_0)$ respectively. $\mathcal{K}$ is the Lie algebra of K, and P is the tangent space of X at 1. $ad(h(e^{2\pi i/8}))$ preserves $\mathcal{P}$, and it gives an endomorphism L satisfying $L^2 = -I$. This makes $\mathcal{P}$ into a complex vector space. Since X is a homogeneous space, all of the tangent spaces become complex vector spaces by translation. This makes X into an almost complex manifold, and in fact a complex manifold. Here is the precise statement.

Theorem 6.3.3. *Suppose that G is a real semisimple algebraic group, and suppose that there is a homomorphism $u : U(1) \to G$ satisfying*

1. *the only characters of $U(1)$ on $Lie(G) \otimes \mathbb{C}$ are 1 and $z^{\pm 1}$.*
2. *$Ad(u(-1))$ is a Cartan involution.*
3. *$u(-1)$ projects nontrivially onto each simple factor of G*

If K is the isotropy group of $ad(u)$ in G^0, then G^0/K has the structure of a homogeneous complex manifold described essentially as above. In particular, G^0 acts by holomorphic automorphisms on it.

A simple algebraic group is one which has no closed normal subgroups other than finite groups. An algebraic group is semisimple if it is "almost" a direct product of simple groups. See Borel [Bo] for the precise definition. Many of the examples we already encountered such as SL_n, SO_n, Sp_{2n}, SU_n, and their products, are all semisimple. A complex manifold of this type is called a *hermitian symmetric space of noncompact type* or simply a hermitian symmetric domain. Let us now suppose that G is the group of real points of an algebraic group over $\mathbb{Q}$ that we also denote by G. Fix faithful representation $G \to GL_n(\mathbb{Q})$. A subroup of $\Gamma \subset G(\mathbb{Q})$ is *arithmetic* if it contains a subgroup $\Gamma_1 \subseteq \Gamma$ such that $\Gamma_1 \subseteq G(\mathbb{Q}) \cap GL_n(\mathbb{Z})$, and such that both inclusions have finite index. Here is the precise statement of a theorem we have used several times already.

Theorem 6.3.4 (Baily-Borel)**.** *Let G, $K \subset G(\mathbb{R})^0$ and $\Gamma \subset H(\mathbb{Q})$ be as above then $\Gamma\backslash G(\mathbb{R})^0/K$ is a quasiprojective variety.*

To apply theorem 6.3.3 in our case, we should replace G by $G^{ad} = G/\pm I$, then we can take "square root" of h, i.e. we can find a character u with $u^2 = h$. To check the first condition of the theorem, it suffices to check that the characters of $U(1)$ on $Lie(G) \otimes \mathbb{C}$ with respect h are 1 and $z^{\pm 2}$.

Lemma 6.3.5. *Suppose that G is as in (6.2), and h as above, then the characters of $U(1)$ on $Lie(G) \otimes \mathbb{C}$ are 1 and $z^{\pm 2}$.*

Proof. We have an inclusion $G(\mathbb{R}) \subseteq Sp_{2g}(\mathbb{R})$, so we may assume equality holds. Let V be the standard representation of $Sp_{2g}(\mathbb{R})$. If $V_\pm$ denote the $\pm i$-eigenspaces of $J_0 = h(i)$ acting on $V \otimes \mathbb{C}$, then $U(1)$ acts on these with character $z^{\pm 1}$ under h. The characters of

$$\mathcal{G} \otimes \mathbb{C} \subseteq V^* \otimes V = End(V_+) \oplus End(V_-) \oplus (V_+^* \otimes V_-) \oplus (V_-^* \otimes V_+)$$

under h are $1, z^{\pm 2}$. □

The second condition holds because $u(-1) = h(i) = J_0$. The third condition is also not hard to check. It follows that X is hermitian symmetric. We take $\Gamma \subset G(\mathbb{Q})$ to the subgroup stabilizing the lattice L. Therefore $\Gamma\backslash X$ is quasiprojective. This is the desired moduli space.

6.4 Deligne's axioms

We briefly indicate Deligne's more abstract axiomatization of Shimura varieties [D1, D2]. Given an algebraic group G, the adjoint group G^{ad} is the image G under its adjoint representation. This has the effect of dividing G by its centre. A connected Shimura datum $\mathcal{S}$ consists a semisimple algebraic group G over $\mathbb{Q}$ a homomorphism $u : U(1) \to G^{ad}(\mathbb{R})$ such that

S1 The characters of $U(1)$ on $Lie(G^{ad}(\mathbb{C}))$ are $z^{-1}, 1, z$.

S2 $Ad\, u(-1)$ is a Cartan involution of $G(\mathbb{C})$.

S3 $G(\mathbb{R})$ is noncompact, and this holds for all $\mathbb{Q}$-factors as well.

These conditions are natural in view of what we said in the previous section. We define $K = \{g \in G(\mathbb{R})^0 \mid gu(x)g^{-1} = u(x)\}$, then $X = G(\mathbb{R})^0/K$ is a hermitian symmetric domain. In fact, we have the following

Proposition 6.4.1. *To give a connected Shimura datum is equivalent to giving*

1. *A semisimple group G of noncompact type (S3 should hold),*
2. *a hermitian symmetric domain X, and*
3. *a surjective homomorphism $G(\mathbb{R})^0 \to Aut(X)^0$ with compact kernel, where $Aut(X)$ is the holomorphic automorphism group of X.*

Proof. [Mi, prop 4.8]. □

Given a connected Shimura datum, and a congruence group $\Gamma \subset G(\mathbb{Q})$, the associated (connected) *Shimura variety* (of finite level) is the quasiprojective variety $Sh(X, \Gamma) = \Gamma \backslash X$. This yields a slick construction of the moduli space of abelian varieties of PEL type, as will now explain. A Shimura datum is said to be of *Hodge type* if there exists an injective homomorphism $\iota : G \hookrightarrow Sp_{2g}$ for some g. The algebraic groups associated to PE data, constructed in the previous section, will have this property. However, the converse is not true. The inclusion ι induces a holomorphic map $\iota' : X \to \mathbb{H}_g$. We can find a congruence group $\Gamma_1 \subset Sp_{2g}(\mathbb{Q})$, containing Γ. The quotient $A'_g = \Gamma_1 \backslash \mathbb{H}_g$ is the moduli space of g-dimensional abelian varieties with and suitable level structure and a fixed polarizational type, which need not be principal. By Borel [Bo2, 3.10], the map $Sh(X, \Gamma) \to A'_g$ induced by ι' is a morphism of algebraic varieties. Thus $Sh(X, \Gamma)$ can be viewed as a moduli space of abelian varieties with extra structure.

In general, Shimura varieties do not admit embeddings into a variety of the form A'_g, and so do not parameterize abelian varieties. Deligne does show that they do parameterize certain Hodge structures, and conjectures a more subtle interpretation involving motives. However, we will not say anything more about this, and to refer to [Mi] for a more detailed introduction to these ideas. Finally, we mention a fact which is very important fact for number theoretic applications.

Theorem 6.4.2. *A Shimura variety is defined over $\bar{\mathbb{Q}}$.*

For A_g, this was Mumford's theorem 4.4.3. The result for Shimura varieties of Hodge type, can be reduced to this. However, in general, it requires a completely different set of ideas. A relatively soft proof for the general case is due to Faltings [F2].

Chapter 7

Siegel modular varieties of small genus

7.1 Genus 2 curves and abelian surfaces

The Siegel upper half plane $\mathbb{H}_2$ is an open subset of the space of 2×2 symmetric matrices. It follows that this, and therefore A_2 is three dimensional. Following Igusa [I1, I2], we will be able give a fairly explicit description of this space. First, we need the following basic result due to Weil [W, Satz 2].

Theorem 7.1.1. *A two dimensional principally polarized variety is isomorphic to either a product of two elliptic curves, or the Jacobian of a smooth projective genus 2 curve.*

Before getting into the proof, we need to make a few general comments. Recall that a principally polarized abelian variety is isomorphic to $A = \mathbb{C}^g/\mathbb{Z}^g + \Omega\mathbb{Z}^g$ for some $\Omega \in \mathbb{H}_g$. Associated to Ω, is the Riemann theta function $\theta(z)$ on $\mathbb{C}^g$. Since this is quasiperiodic, the zero set of θ gives a well defined effective divisor Θ on A. This is the geometric "incarnation" of the theta function. We have three facts, whose proofs can be found in [BL, MAV]:

1. Θ is ample.
2. $\dim H^0(\mathcal{O}_A(\Theta)) = 1$.
3. $H^i(\mathcal{O}_A(\Theta)) = 0$ for $i > 0$

We should point out that 2. depends on the fact that our polarization is principal. 3. is really just a special case of Kodaira's vanishing theorem, but the case of abelian varieties is a lot easier. Let us now return to the case at hand, where $g = 2$. Recall that we have an intersection pairing for divisors on a surface, so in particular on A. The proof theorem 7.1.1 can be broken down into a series of lemmas.

Lemma 7.1.2. *Suppose that $g = 2$, then $\Theta^2 = 2$.*

Proof. Combining the above the above facts with Riemann-Roch 3.4.3 yields

$$1 = \chi(\mathcal{O}_A(\Theta)) = \frac{1}{2}\Theta^2$$

□

Lemma 7.1.3. *If $C \subset A$ is an irreducible curve, $C^2 \geq 0$.*

Proof. We use the fact that $C^2 = C\cdot C'$ for any curve C' algebraically equivalent to C. See [H, p 367] for an explanation of algebraic equivalence, and for the fact just stated. If C' is a translate of C by a nonzero element of A, then it follows that $C^2 = C \cdot C'$. On the other hand, $C \cdot C' \geq 0$ by the formula (3.3). □

Lemma 7.1.4. *If $C_1, C_2 \subset A$ are two irreducible curves with C_1 an elliptic curve, then either they are algebraically equivalent or $C_1 \cdot C_2 > 0$.*

Proof. The quotient A/C_1 has the structure of an elliptic curve, and the projection $\pi : A \to A/C_1$ is holomorphic. The image of C_2 under the π is either a point or all of A/C_1. In the first case, C_2 is fibre of π, so it is algebraically equivalent to C_1. In the second case, C_1 and C_2 must intersect, so $C_1 \cdot C_2 > 0$. □

Lemma 7.1.5. *Either Θ is a smooth curve of genus 2, or $\Theta = C_1 + C_2$ where C_i are elliptic curves and $C_1 \cdot C_2 = 1$.*

Proof. Let us suppose that $\Theta = \sum C_i$, where C_i are irreducible curves with repetitions allowed. Then

$$2 = \Theta^2 = \sum C_i^2 + 2\sum C_i \cdot C_j \tag{7.1}$$

All the terms in the sums are nonnegative, so this puts strong constraints on what we can have. Either we have

(a) a single curve $\Theta = C = C_1$ with $C^2 = 2$,

(b) we have two curves satisfying $C_1^2 = C_2^2 = 0$ and $C_1 \cdot C_2 = 1$,

(c) or more than two curves.

The rest of the proof will hinge on the adjunction formula [H, p 361, p 366] that if C is a possibly singular curve with arithmetic genus h on a surface S, then

$$2h - 2 = C \cdot (K + C)$$

Applied to $S = A$, this says

$$2h - 2 = C^2$$

Let us suppose we are case (a). So we have an curve $C = \Theta$ with $C^2 = 2$. Then C must have arithmetic genus 2. If it is smooth then it has genus 2. Now suppose that C is singular. Let $\tilde{C} \to C$ denote the normalization. The genus $\tilde{h}$

is known to be strictly smaller than the arithmetic genus 2. So either $\tilde{h} = 0$ or 1. In the first case, we would get a nonconstant map from $\mathbb{P}^1$ to A, which would lift to a nonconstant map from $\mathbb{P}^1 \to \mathbb{C}^2$. This would contradict Liouville's theorem. In the second case, $\tilde{C} \to A$ is the a homomorphism up to translation by [BL, prop 1.2.1]. This would mean that C is nonsingular, contrary to what we assumed.

In case (b), one of curves, say C_1 has $C_1^2 = 0$, otherwise (7.1) will fail. The adjunction formula, shows that C_1 has arithmetic genus 1. Arguing as above, one sees that C_1 is nonsingular, and therefore an elliptic curve. By the previous lemma, either C_1 and C_2 are algebraically equivalent, or $C_1 \cdot C_2 > 0$. If C_i are equivalent, then $(C_1 + C_2)^2 = 0$ which contradicts (7.1). So $C_1 \cdot C_2 > 0$. In this case, we must have $C_1 \cdot C_2 = 1$ and $C_2^2 = 0$. Therefore C_2 is also an elliptic curve.

We claim that the final case (c) cannot occur. For simplicity, let's assume we have exactly three curves C_i. As above, we can assume $C_1^2 = 0$ and that it is elliptic, and either C_1, C_2 are equivalent, or that $C_1 \cdot C_2 = 1$. In the first case, $\Theta = 2C_1 + C_3$ (up to algebraic equivalence). Then Θ^2 would be either 0 or at least 4, but this contradicts (7.1). So we can conclude that $C_1 \cdot C_2 = 1$. But by the same argument, we must also have $C_1 \cdot C_3 = 1$. So that $\Theta^2 \geq 4$, which is again a contradiction. □

Lemma 7.1.6. *If $\Theta = C_1 + C_2$, where C_i are elliptic curves satisfying $C_1 \cdot C_2 = 1$, then A is isomorphic to a product of elliptic curves.*

Proof. By [H, chap V, lemma 1.3], $\deg \Theta|_{C_1} = C_1 \cdot (C_1 + C_2)$. The adjunction formula implies that $C_1^2 = 0$, so $\deg \Theta|_{C_1} = 1$. This means that Θ restricts to a principal polarization on C_1. Now use remark 5.2.5. (With a bit more work, we can see that $A \cong C_1 \times C_2$.) □

The Jacobian of a curve can be generalized as follows [GH, p 331]. Given a smooth projective variety X, the Albanese is the quotient

$$Alb(X) = \frac{H^0(X, \Omega_X^1)^*}{H_1(X, \mathbb{Z})}$$

It is in fact an abelian variety. Furthermore, given the Abel-Jacobi can be defined exactly as before to give a holomorphic map

$$AJ : X \to Alb(X)$$

with $AJ(x_0) = 0$. It is not difficult to show that $(Alb(X), AJ)$ is universal way to map X into an abelian variety. It is easy to see that the construction is functorial in the sense that a holomorphic $f : X \to Y$ induces a homomorphism $Alb(X) \to Alb(Y)$, and furthermore that if $f_* : H_1(X, \mathbb{Z}) \to H_1(Y, \mathbb{Z})$ is an isomorphism, then so is the map on Albaneses.

Lemma 7.1.7. *If $\Theta = C$ is a nonsingular genus two curve, then $A \cong J(C)$.*

Proof. The Lefschetz hyperplane theorem [Ml] implies that the natural map $H_1(C,\mathbb{Z}) \to H_1(A,\mathbb{Z})$ is surjective. It must therefore be an isomorphism because they are both free of the same rank. So by the above remarks, we have an isomorphism

$$J(C) = Alb(C) \cong Alb(A) \cong A$$

□

7.2 Explicit models for A_2

Let us start by describing genus two curves. Given a degree 6 polynomial $f(x)$, with distinct roots, the smooth projective curve determined by

$$z^2 = f(x)$$

has genus 2. This follows from the Riemann-Hurwitz formula. Conversely, any genus 2 can be realized this way (see for example [H, p 304]), but the choice of f is far from unique. The nonuniqueness can be understood. First replace f by the homogenous polynomial $F(x,y)$, then $SL_2(\mathbb{C})$ acts on the space V of such polynomials through the standard action on $\begin{pmatrix} x \\ y \end{pmatrix}$. We are interested in invariant polynomials $V \to \mathbb{C}$ of even degree. These form a ring whose generators were known since the 19th century by Clebsch and Bolza. To describe them, let us factor

$$f(x) = u_0x^6 + u_1x^5 + \dots u_6 = u_0 \prod (x - r_i)$$

or equivalently

$$F(x,y) = u_0x^6 + u_1x^5y + \dots u_6y^6 = \prod (p_ix - q_iy)$$

The standard action of $SL_2(\mathbb{C})$ on the $\begin{pmatrix} p_i \\ q_i \end{pmatrix}$ is compatible with the action on F. The expressions

$$R_{ij} = \det \begin{pmatrix} p_i & p_j \\ q_i & q_j \end{pmatrix}^2$$

are clearly invariant under $SL_2(\mathbb{C})$. We also introduce

$$A(u) = R_{12}R_{34}R_{56} + \sum_{\sigma \in S_6 - \{1\}} R_{\sigma(1)\sigma(2)}R_{\sigma(3)\sigma(4)}R_{\sigma(5)\sigma(6)}$$

$$B(u) = R_{12}R_{23}R_{31}R_{45}R_{56}R_{64} + \dots$$

$$C(u) = R_{12}R_{23}R_{31}R_{45}R_{56}R_{64}R_{14}R_{25}R_{36} + \dots$$

$$D(u) = \prod_{i<j} R_{ij}$$

where the last two sums are symmetrized in the same way. Note that by the theorem on elementary symmetric functions, A, B, C, D are polynomials in the

u's. These are also invariant, because the expressions R_{ij} are invariant. The expressions $A(u), B(u), C(u), D(u)$ generate the ring of invariants. Note that

$$D(u) = u_0^{10} \prod_{i<j} (r_i - r_j)^2$$

is just the discriminant, so that $D(u) \neq 0$ if and only if f has distinct roots. A genus 2 curve corresponds to a $PGL_2(\mathbb{C})$ orbit of the point $[u] \in \mathbb{P}(V)$ with $D(u) \neq 0$. This is determined by the ratios of the values of the fundamental invariants. Here is the precise statement:

Theorem 7.2.1 (Igusa)**.** *The complement of the divisor $D = 0$ in the projective variety $\operatorname{Proj} \mathbb{C}[A, B, C, D]$ is the moduli space M_2 of genus 2 curves. Furthermore, M_2 is isomorphic to $\mathbb{C}^3/\mu_5$, where the group of 5th roots of unity μ_5 acts by $(t_1, t_2, t_3) \mapsto (\zeta t_1, \zeta^2 t_2, \zeta^3 t_3)$.*

While the two descriptions of M_2 may appear unconnected, he derived the second from the first by an explicit set of transformations. Define new invariants by

$$\begin{aligned}
J_2 &= \frac{1}{8} A \\
J_4 &= \frac{1}{96}(4J_2^2 - B) \\
J_6 &= \frac{1}{576}(8J_2^3 - 160 J_2 J_4 - C) \\
J_{10} &= 2^{-12} D
\end{aligned}$$

Then Igusa [I1, I2] showed that the degree 0 part of the graded ring $\mathbb{C}[A, B, C, D^{\pm 1}] = \mathbb{C}[J_2 \ldots, J_{10}^{\pm 1}]$ is isomorphic to $\mathbb{C}[t_1, t_2, t_3]^{\mu_5}$, via

$$J_2^{e_1} J_4^{e_2} J_6^{e_3} J_{10}^{-e_5} \mapsto t_1^{e_1} t_2^{e_2} t_4^{e_3}$$

where

$$e_1 + 2e_2 + 3e_3 = 5e_5$$

We have an injective map $\mathbb{H}^2 \to \mathbb{H}_2$ given by

$$(\tau_1, \tau_2) \mapsto \begin{pmatrix} \tau_1 & 0 \\ 0 & \tau_2 \end{pmatrix}$$

Let $\Delta \subset A_2$ denote the image of $\mathbb{H}^2$. This is a divisor parameterizing products of pairs of elliptic curves. From theorem 7.1.1, we see that the complement consists of Jacobians of genus 2 curves. Given a curve, $J(C)$ with its canonical principal polarization, the curve can be recovered simply as the theta divisor Θ. Therefore, we have proved:

Theorem 7.2.2. *A_2 contains a divisor Δ parameterizing products of pairs of elliptic curves. The complement of $A_2 - \Delta \cong M_2$.*

A Siegel modular form of weight k is a holomorphic function f on $\mathbb{H}_2$ which transforms in the "expected way" under $Sp_4(\mathbb{Z})$:

$$f((\alpha\Omega+\beta)(\gamma\Omega+\delta)^{-1}) = \det(\gamma\Omega+\delta)^k f(\Omega), \quad \begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} \in Sp_4(\mathbb{Z})$$

One important class of examples are Eisenstein series, which are sums of the form

$$E_k(\Omega) = \sum \det(\gamma\Omega+\delta)^{-k}$$

where sum runs over suitable set of pairs of matrices (γ, δ). The sum of the spaces of modular forms over all weights forms a graded algebra. As an application of the last theorem, Igusa [I2] goes on to determine an explicit set of generators and relations for this algebra. See also [G3, §9].

Theorem 7.2.3 (Igusa)**.** *The algebra of Siegel modular forms of even weight is generated by the Eisenstein series E_4, E_6, E_{10} and E_{12}.*

Let us consider the moduli space $A_{2,3}$ of abelian surfaces with level three structure. There is an explicit birational model which can be described as follows. Let σ_i denote the ith elementary symmetric polynomial in $x_0, \ldots, x_5$. Burkhardt's quartic B is the variety in $\mathbb{P}^5$ defined by $\sigma_1 = \sigma_4 = 0$. Since $\sigma_1 = 0$ is linear, one of the variables can be eliminated so as to write B as a quartic hypersurface in $\mathbb{P}^4$. This can be transformed to yield a more economical representation

$$b(y_0, \ldots, y_4) = y_0(y_0^3 + y_1^3 + y_2^3 + y_3^3 + y_4^3) + 3y_1y_2y_3y_4 = 0$$

[Hu, chap 5]. The variety B has 45 nodes (isolated singularities analytically isomorphic to $x_1^2 + \ldots x_4^2 = 0$) but no other singularities. Let $\tilde{B}$ be resolution of singularities of B given by the blow up of B at the nodes. The following fact was known in some form by Burkhardt in the 19th century (see [HW, p 3270] for the history). A modern proof can be found in [G1].

Theorem 7.2.4. *There exists a birational map $A_{2,3} \to B$ given by an explicit set of Siegel modular forms. This extends to an open immersion $A_{2,3} \hookrightarrow \tilde{B}$.*

Finally, to relate this to what we did earlier, recall that Hilbert modular surfaces are moduli spaces of abelian surfaces with multiplication by a real quadratic field. Therefore they map to A_2, and to $A_{2,n}$ under appropriate conditions. We can explicitly identify one of these surfaces using the previous model.

Theorem 7.2.5 (Van der Geer-Hirzebruch)**.** *The image of the Hilbert modular surface in B corresponding to $\mathbb{Q}(\sqrt{5})$ and the level 3 principal Hilbert modular group is given by $\sigma_1 = \sigma_2 = \sigma_4 = 0$.*

Proof. [G1]. ☐

7.3 Compactification of A_2

The spaces A_g are noncompact. Fortunately, there is a natural compactification A_g^* given by Baily-Borel [BB]. In this case, the basic ideas go back to Satake [Sa], so it also called the Satake compactification. To describe it in more explicit terms, it is convenient to replace $\mathbb{H}_g$ by

$$D_g = \{Z \in Mat_{g\times g}(\mathbb{C}) \mid Z^T = Z, I - Z\bar{Z} \text{ pos. def.}\}$$

There is a holomorphic isomorphism $\mathbb{H}_g \xrightarrow{\sim} D_g$ given by the Cayley transform

$$\Omega \mapsto (\Omega - iI)(\Omega + iI)^{-1}$$

D_g is a bounded domain in $\mathbb{C}^{g^2}$. The group $Sp_{2g}(\mathbb{R})$ will act on D_g, and on its closure via the previous isomorphism. For simplicity, let us now assume $g = 2$. Then

$$\{I\}, \{\begin{pmatrix} \tau & 0 \\ 0 & 1 \end{pmatrix} \mid \tau \in D_1\} \subset \partial D_2$$

Let D_2^* be the union of D_2, the above sets, and their translates under the group $Sp_{2g}(\mathbb{Q})$.

Theorem 7.3.1 (Satake)**.** *There is a Hausdorff topology on D_2^*, extending the usual one on D_2, such that $Sp_4(\mathbb{Z})$ acts properly discontinuously, and such that $A_2^* = Sp_4(\mathbb{Z})\backslash D_2^*$ is compact. As a set, we can decompose A_2^* as a disjoint union*

$$A_2^* = A_2 \cup A_1 \cup A_0$$

By work of Baily-Borel, A_2^* has the structure of a projective algebraic variety. However, the singularities are quite bad. Igusa was able to construct a different compactification $\bar{A}_2$, with mild singularities, by explicitly blowing up A_2^*. This turns out to coincide with the so called Deligne-Mumford compactification of M_2 [DM]. As a bonus, this gives a moduli interpretation for points at infinity. A possibly reducible projective curve is called *stable* if it is reduced and connected, with at worst nodal singularities, and with finite automorphism group. The last condition is equivalent to requiring that any $\mathbb{P}^1$ component meets the other components in at least 3 points.

Theorem 7.3.2 (Deligne-Mumford)**.** *The points of $\bar{A}_2$ correspond to stable curves with arithmetic genus two.*

We can write

$$\bar{A}_2 = \underbrace{M_2 \cup \Delta_1}_{A_2} \cup \Delta_0$$

where Δ_i are divisors.The general points of Δ_1 correspond to unions of two elliptic curves meeting transversally, and Δ_0 to irreducible curves with a single node. These constructions can be applied to varieties with level structure. In particular $\bar{A}_{2,3}$ is the variety $\tilde{B}$, constructed previously, as the blow up of the

Burkhardt quartic. Let $M_{2,3}$ be the preimage of M_2 in $A_{2,3}$. Its image in B is given by the nonvanishing of the Hessian

$$\det\left(\frac{\partial^2 b}{\partial y_i \partial y_j}\right) \neq 0,$$

and the universal genus 2 curve over it is described by an explicit equation in the recent paper of Bruin and Nasserden [BN].

7.4 Genus 3 curves

Let us take quick a look at the next case. An easy dimension count shows that $\dim A_3 = 6$. The first step works as before:

Theorem 7.4.1 (Oort-Ueno)**.** *A principally polarized abelian three dimensional variety is either the Jacobian of a genus 3 curve, a product of the Jacobian of a genus 2 curve and an elliptic curve, or a product of three elliptic curves.*

Proof. [OU]. □

The products are contained in the image of $\mathbb{H}_2 \times \mathbb{H}$, and this can be seen to be a proper Zariski closed set of A_3. Therefore the complement M_3, parameterizing Jacobians of genus 3 curves, is a nonempty Zariski open (and therefore dense) subset of A_3. We have the following important fact.

Theorem 7.4.2 (Torelli)**.** *If X is a smooth projective curve, $(J(X), \Theta)$ determines X up to isomorphism.*

Proof. [GH, p 359]. □

Corollary 7.4.3. *M_3 can identified with the set of isomorphism classes of genus 3 curves.*

So we now focus on genus 3 curves. Our goal will be to give a direct construction of a part of M_3 as an algebraic variety. Suppose that X is a genus 3 curve. Choose a basis $\omega_0, \omega_1, \omega_2$ of the space holomorphic 1-forms. Then we get the so called canonical map $X \to \mathbb{P}^2$ given by $x \mapsto [\omega_0(x), \omega_1(x), \omega_2(x)]$, roughly speaking.

Proposition 7.4.4. *The canonical map is either a degree two map onto a conic, or an isomorphism to a quartic in $\mathbb{P}^2$. Any nonsingular quartic in $\mathbb{P}^2$ arises this way.*

Proof. [H, pp 341-342]. □

Both cases are possible, but the second is typical in the sense that it holds on an open dense subset $M_3^o \subset M_3$ of the moduli space. Our goal is to construct M_3^o from first principles. Let us first describe the space of nonsingular quartic curves. A homogeneous quartic polynomial $f(x, y, z) = u_0 x^4 + \ldots u_{14} z^4$ has

15 coefficients. Thus it forms a 15 dimensional vector space V. The space of quartics curves is parameterized by $\mathbb{P}^{14} = \mathbb{P}(V)$. Using basic algebraic geometry, we can find a polynomial $D(u)$, called the discriminant, such that $D(u) = 0$ if and only if $V(f)$ is a singular curve [Mk, p 170]. Therefore the affine variety $Q = \mathbb{P}^{14} - V(D)$ parameterizes nonsingular quartics. We have a map $\pi : Q \to M_3^o$, but this is not a bijection. An element of Q is a genus 3 curve plus an embedding into the plane. The map π forgets the embedding. We note that $GL_3(\mathbb{C})$ acts on the space of polynomials V, and so it induces an action of $PGL_3(\mathbb{C})$ on $\mathbb{P}^{14}$. The set Q is invariant under this action. Two points correspond to isomorphic curves if and only if they lie on the same orbit correspond. Thus π is the quotient map. We want to show that $M_3^o = Q/PGL_3(\mathbb{C})$ is an algebraic variety. The name of the game is *geometric invariant theory*, which was invent by Mumford [GIT] precisely for the purpose of constructing moduli spaces. Here is the general result in the form that we need:

Theorem 7.4.5 (Hilbert, Mumford, Nagata)**.** *Suppose that G is a reductive linear algebraic group acting on an affine variety $X = \operatorname{Spec} R$. Then the ring of invariants R^G is finitely generated; $X//G := \operatorname{Spec} R^G$ is called the GIT quotient. There is a surjective map of sets $X/G \to X//G$, which is a bijection if all points of X have finite isotropy group.*

Proof. See [Mk, pp 135-137, 165-167] □

Note that $\operatorname{Spec} R$ is tacitly taken to be the maximal ideal spectrum. We can take "reductive" to mean that there is a compact Zariski dense subgroup $K \subset G$. For example, $GL_n(\mathbb{C})$ (resp. $PGL_n(\mathbb{C})$) is reductive because $K = U_n$ (resp. $\operatorname{im} U_n$) is such a subgroup. Without the reductivity hypothesis, R^G can fail to be finitely generated. The map $X/G \to X//G$ need not be bijective; when it is, we say that $X//G$ is the geometric quotient.

Now returning to our original problem, we can form the GIT quotient $Q//PGL_3(\mathbb{C})$, and verify that it is a geometric quotient. This will construct M_3^o. We have to check that the isotropy groups are finite, but this follows from [Mk, thm 5.23]; we could also argue that the isotropy group of a point corresponding to a curve X is $Aut(X)$, but this well known to be finite.

As a consequence of this construction, we find that

Proposition 7.4.6. *M_3 (and therefore) A_3 is rationally connected, i.e. two general points can be connected by a rational curve.*

Proof. There is a dominant rational map from $\mathbb{P}^{14}$ to M_3. Therefore two general points can be connected by the image of a line. □

Bibliography

[BB] W. Baily, A. Borel: *Compactification of arithmetic quotients of bounded symmetric domains.* Ann. of Math, 84 (1966)

[BPV] W. Barth, C. Peters, Van de Ven, *Compact complex surfaces*, Springer (1984)

[BL] C. Birkenhake, H. Lange, *Complex abelian varieties*, Springer (2003)

[Bo] A. Borel, *Linear algebraic groups*, Benjamin (1969)

[Bo2] A. Borel, *Some metric properties of arithmetic quotients of symmetric spaces and an extension theorem*, J. Diff. Geom (1972)

[BN] N. Bruin, B. Nasserden, *Arithmetic aspects of the Burkhardt quartic threefold.* J. Lond. Math. Soc. 98 (2018)

[C] H. Cartan et. al., *Sem. Henri Cartan, vol 10* 1957-58

[DM] P. Deligne, D. Mumford,*The irreducibility of the space of curves of given genus.* IHES No. 36 1969 75-109.

[D1] P. Deligne, *Travaux de Shimura*, Sem. Bourbaki (1971)

[D2] P. Deligne, *Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques*, Automorphic forms, representations and L-functions (1979)

[DR] P. Deligne, M. Rapoport, *Les schémas de modules de courbes elliptiques*, SLN 349, Springer (1973)

[DMOS] P. Deligne, J. Milne, A. Ogus, K. Shih, *Hodge cycles, motives and Shimura varieties*, Springer (1982)

[DM] H. Dym, H. McKean, *Fourier series and integrals*, Academic Press (1972)

[DS] F. Diamond, J. Shurman, *A first course in modular forms*, Springer (2005)

[F1] G. Faltings, *Endlichkeitssätze für abelsche Varietten über Zahlkörpern.* Invent. Math. (1983)

[F2] G. Faltings, *Arithmetic varieties and rigidity,* Sem. on number theory, Birkäuser (1984), 63-77

[F] E. Freitag, *Hilbert modular forms*, Springer (1990)

[G1] G. van der Geer, *Note on abelian schemes of level three*, Math. Ann. (1987)

[G2] G. van der Geer, *Hilbert modular surfaces*, Springer (1988)

[G3] G. van der Geer, *Siegel modular forms and their applications*, The 1-2-3 of modular forms, 181-245, Springer, (2008)

[GR] H. Grauert,R. Remmert, *Coherent Analytic Sheaves*, Springer (1984)

[GH] P. Griffiths, J. Harris, *Principles of algebraic geometry*, Wiley (1978)

[H] R. Hartshorne, *Algebraic geometry*, Springer (1977)

[HW] W. Hoffman, S. Weintraub, *The Siegel modular variety of degree two and level three*, Trans. AMS 2000

[Hu] B. Hunt, *The geometry of some special arithmetic quotients*, Springer (1996)

[I1] J. Igusa *Arithmetic variety of moduli for genus two.* Ann. of Math. 72 (1960)

[I2] J. Igusa, *On Siegel modular forms of genus two.* Amer. J. Math. 84 (1962)

[LR] R. Langlands, D. Ramakrishnan, (ed.) *The zeta functions of Picard modular surfaces* CRM (1992)

[Mi] J. Milne, *Introduction to Shimura varieties*, Harmonic analysis, the trace formula, and Shimura varieties, AMS (2005)

[Ml] J. Milnor, *Morse theory*, Princeton (1959)

[Mk] S. Mukai, *An introduction to invariants and moduli*, Cambridge (2003)

[GIT] D. Mumford, *Geometric invariant theory*, Springer (1965)

[MAV] D. Mumford, *Abelian varieties*, Tata Inst. (1970)

[MT] D. Mumford, *Tata lectures on theta I*, Birkauser (1983)

[M] D. Mumford, *A Note on Shimura's paper discontinuous groups and abelian varieties*, Math Ann (1969)

[OU] F. Oort,, K. Ueno, *Principally polarized abelian varieties of dimension two or three are Jacobian varieties.* J. Fac. Sci. Univ. Tokyo Sect. IA Math. 20 (1973), 377-381.

[GAGA] J.P. Serre, *Géométrie algébrique et géométrie analytic*, Ann. Inst. Fourier (1955)

[Rig] J.P. Serre, *Rigiditeé du foncteur de Jacobi d' échelon* $n \geq 3$. Sem. Cartan (1960)

[Sa] I. Satake, *On the compactification of the Siegel space*, J. Indian Math. Soc. , 20 (1956) pp. 259281

[Se] J.P. Serre, *A course in arithmetic*, Springer (1973)

[Si] J. Silverman, *Arithmetic of elliptic curves*, Springer (1986)

[S1] G. Shimura, *On families of polarized abelian varieties and automorphic functions*, Ann Math (1963)

[S2] G. Shimura, *Arithmetic theory of automorphic functions*, Princeton (1971)

[W] A. Weil, *Zum Beweis des Torellischen Satzes,* Nachr. Akad. Wiss. Göttingen, Math.-Phys. (1957)