CHAPTER 4

Heegner points on Shimura curves

Any elliptic curve E over $\mathbb{Q}$ is modular, and hence is equipped with the modular parametrisation

$$\Phi_N : \mathcal{H}/\Gamma_0(N) \longrightarrow E(\mathbb{C}), \quad \text{where } N = \text{ conductor of } E, \tag{4.1}$$

as introduced in Chapter 2. The theory of complex multiplication of Chapter 3 allows the construction of a plentiful supply of algebraic points on E—the so-called Heegner points, of the form $\Phi_N(\tau)$, where $\tau \in \mathcal{H}$ is a quadratic (imaginary) irrationality.

In particular, if K is an imaginary quadratic field satisfying the Heegner hypothesis, then for all orders $\mathcal{O}$ of K of conductor prime to N, the set $CM(\mathcal{O})$ of points in $\mathcal{H}/\Gamma_0(N)$ with associated order equal to $\mathcal{O}$ is non-empty, and it is possible to choose points $\tau_n \in CM(\mathcal{O}_n)$ in such a way that the collection of points $P_n = \Phi_N(\tau_n)$ forms a *Heegner system* in the sense of Definition 3.12 of Chapter 3. This Heegner system is an essential ingredient in the proof of the theorem of Gross-Zagier-Kolyvagin stated in Chapter 1.

It is natural to examine what happens if the Heegner hypothesis is relaxed. For example, suppose that $N = p$ is a prime which is *inert* in K. One can show (cf. Exercise 1) that if τ belongs to $\mathcal{H} \cap K$, then $P_\tau := \Phi_N(\tau)$ belongs to $E(H_n)$ for some n of the form $p^t n'$ with $t \geq 1$ and $(p, n') = 1$. Furthermore,

$$\mathrm{Trace}_{H_n/H_{n'}}(P_\tau) = 0.$$

Thus the Heegner point construction does not yield any points on E defined over ring class fields of conductor prime to p. This is to be expected, since $S_{E,K} = \{p, \infty\}$ so that $\mathrm{sign}(E, K)$ is equal to 1: in this case, one expects the rank of $E(H_{n'})$ to be small in general.

A second example which is more interesting, and which the reader may find helpful to keep in mind in a first reading of this chapter, is the one where $N = pq$ is a product of two distinct primes p and q which are both inert in $K/\mathbb{Q}$. In that case, $S_{E,K} = \{p, q, \infty\}$, so that $\mathrm{sign}(E, K) = -1$. As in the previous example, the points of the form $\Phi_N(\tau)$ belong to $E(H_n)$ where n is of the form $p^r q^s n'$ with $r, s \geq 1$, and the trace of these points to $E(H_{p^r n'})$ or to $E(H_{q^s n'})$ are torsion. It thus appears that the modular parametrisation Φ_N is inadequate to produce the non-trivial Heegner system whose existence is predicted by Conjecture 3.16.

To deal with this example and its obvious generalisations, it seems essential to enlarge the répertoire of modular parametrisations to include *Shimura curve* parametrisations as well as the more classical modular curve parametrisation of (4.1).

4.1. Quaternion algebras

A quaternion algebra over a field F is a 4-dimensional central simple algebra over F. A trivial example is the ring $M_2(F)$ of 2×2 matrices with entries in F. Any quaternion algebra over a field F of characeristic $\neq 2$ is isomorphic to an algebra of the form

$$\left(\frac{a,b}{F}\right) := F \oplus Fi \oplus Fj \oplus Fk, \quad \text{where } i^2 = a, j^2 = b, ij = -ji = k, \tag{4.2}$$

for some $a, b \in F^\times$. A quaternion algebra B over F is said to be *split* if it is isomorphic to $M_2(F)$. More generally, if K is an extension field of F, then B is said to be split over K if $B \otimes_F K$ is a split quaternion algebra over K.

Every quaternion algebra splits over some extension of F (for example, any maximal commutative subfield of B). There are, up to isomorphism, exactly two quaternion algebras over the reals: the split algebra $M_2(\mathbb{R})$ and the algebra $\mathbb{H}$ of Hamilton's quaternions. A similar fact is true over $\mathbb{Q}_p$ or any local field. All of this is elementary. (Cf. Exercise 2.)

More deep is the classification of quaternion algebras over number fields, which, together with the more general classification of central simple algebras, is a cornerstone of global class field theory. (Cf. [**CF67**].) For any place v of F, let F_v denote the completion of F at v and let $B_v := B \otimes_F F_v$. One says that B is *split at* v if B_v is a split quaternion algebra. Otherwise B is said to be *ramified* at v.

PROPOSITION 4.1. *Let S be a finite set of places of $\mathbb{Q}$. Then there exists a quaternion algebra ramified precisely at the places in S, if and only if S has even cardinality. In this case the quaternion algebra is unique up to isomorphism.*

Let Z be a finitely generated subring of F. (Of principal interest are the cases where $Z = \mathcal{O}_F$ is the ring of integers of F, or where Z is the ring of S-integers for some finite set S of places of F.)

DEFINITION 4.2. A *Z-order* in B is a subring of B which is free of rank 4 as a Z-module. A *maximal Z-order* is a Z-order which is properly contained in no larger Z-order. An *Eichler Z-order* is the intersection of two maximal Z-orders.

The *level* of an Eichler order $R = R_1 \cap R_2$ is the Z-module index of R in either R_1 or R_2. One can show (cf. Exercise 5) that this notion is independent of the description of R as an intersection of two maximal orders.

Unlike the rings of integers of number fields of which they are the non-commutative counterpart, maximal Z-orders in a quaternion algebra are never unique. This is because any conjugate of a maximal order is also a maximal order. The most one can ask for in general is that a maximal Z-order be unique up to conjugation by elements of $B^\times$. Such uniqueness is not true in general, but it is under the following general condition:

DEFINITION 4.3. One says that that B and Z satisfy the *Eichler condition* if there is at least one archimedean prime or one prime which is invertible in Z at which B is split.

PROPOSITION 4.4. *Suppose that B and Z satisfy the Eichler condition. Then any two maximal Z-orders in B are conjugate. Likewise, any two Eichler Z-orders of the same level are conjugate.*

The proof of this proposition is explained in [**Vi80**]. More precisely, ch. III, §5, of [**Vi80**] describes the set of Eichler Z-orders of a given level N in terms of an adelic double coset space attached to B. To make this explicit, let $\hat{\mathbb{Z}}$ denote the usual profinite completion of $\mathbb{Z}$ and write $\hat{\mathbb{Q}} := \hat{\mathbb{Z}} \otimes \mathbb{Q}$ for the ring of finite rational adèles. Fixing one Eichler Z-order R of level N in B, let

$$\hat{R} := R \otimes \hat{\mathbb{Z}}, \qquad \hat{B} := B \otimes \hat{\mathbb{Q}} = \hat{R} \otimes \mathbb{Q}$$

denote the "adelisations" of R and B respectively. Then the set of Eichler Z-orders of level N in B is in natural correspondence with the coset space

$$\hat{B}^\times / \hat{\mathbb{Q}}^\times \hat{R}^\times,$$

by assigning to the coset represented by an idèle (b_ℓ) (indexed by rational primes ℓ) the order

$$(b_\ell)\hat{R}(b_\ell^{-1}) \cap B.$$

It can be checked that this an Eichler Z-order in B of level N which depends only on the coset of (b_ℓ) and not on the choice of a representative, and that all Eichler Z-orders in B of level N are obtained in this way. It follows that the set of conjugacy classes of Eichler Z-orders of level N in B is in natural bijection with the double coset space

$$B^\times \backslash \hat{B}^\times / \hat{R}^\times. \tag{4.3}$$

Given any rational prime p, let $B_p := B \otimes \mathbb{Q}_p$ and let $R_p := R \otimes \mathbb{Z}_p$. The following *strong approximation theorem* yields a p-adic description of the double coset space appearing in (4.3):

THEOREM 4.5. *Let p be a prime at which the quaternion algebra B is split. Then the natural map*

$$R[1/p]^\times \backslash B_p^\times / R_p^\times \longrightarrow B^\times \backslash \hat{B}^\times / \hat{R}^\times,$$

which sends the class represented by b_p to the class of the idèle $(\ldots, 1, b_p, 1, \ldots)$, is a bijection.

For further discussion see ch. III, §4 of [**Vi80**] or Section 0.2 of [**Cl03**].

Any quaternion algebra B over F admits a natural four-dimensional linear representation over F by letting B act on itself by left multiplication. Given $b \in B$, the corresponding F-linear endomorphism of B has a characteristic polynomial of the form

$$f_b(x) = (x^2 - tx + n)^2.$$

The integers t and n are called the *reduced trace* and the *reduced norm* of x respectively. (See Exercise 3 and [**Vi80**] for more details.)

4.2. Modular forms on quaternion algebras

Let B be a quaternion algebra over $\mathbb{Q}$ which is split at ∞. (Such an algebra is called an *indefinite* quaternion algebra.) Fix an identification

$$\iota : B \otimes_{\mathbb{Q}} \mathbb{R} \simeq M_2(\mathbb{R}).$$

Let R be an order in B. Denote by $R_1^\times$ the group of elements of $R^\times$ of reduced norm 1, and let

$$\Gamma := \iota(R_1^\times) \subset \mathbf{SL}_2(\mathbb{R}).$$

LEMMA 4.6. *The group Γ acts discretely on $\mathcal{H}$ with compact quotient.*

PROOF. Since R is discrete in $B \otimes \mathbb{R}$, the group $R_1^\times$ is discrete in $(B \otimes \mathbb{R})^\times$, so that Γ is a discrete subgroup of $\mathbf{SL}_2(\mathbb{R})$. But $\mathcal{H}$ is identified with the coset space $\mathbf{SL}_2(\mathbb{R})/\mathbf{SO}_2(\mathbb{R})$ where $\mathbf{SO}_2(\mathbb{R})$ is the stabiliser of i. Since this latter group is compact, the discreteness of the action of Γ on $\mathcal{H}$ follows. The proof that the action of Γ on $\mathcal{H}$ has a fundamental region with compact closure, which uses in an essential way the assumption that Γ arises from a quaternion division algebra, can be found for example in [**Ka92**], thm. 5.4.1. □

DEFINITION 4.7. A modular form of weight k on Γ is a holomorphic function f on $\mathcal{H}$ such that

$$f(\gamma\tau) = (c\tau + d)^k f(\tau) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma.$$

REMARK 4.8. It is not necessary to assume any growth conditions, since the quotient $\mathcal{H}/\Gamma$ is already compact. In this sense the theory of modular forms attached to non-split quaternion algebras is simpler than the classical theory of forms on $\Gamma_0(N)$. We will see shortly that the absence of cusps is also a source of extra difficulties in the theory, since the notion of Fourier expansions at the cusps is used crucially in the proof of the multiplicity one theorem of Lemma 2.7, in the integral representation of $L(f, s)$, and in Proposition 2.11 giving an explicit formula for the modular parametrisation Φ_N.

As in the classical setting where $B = M_2(\mathbb{Q})$, the main case which is relevant for elliptic curves and modular parametrisations is the one where $k = 2$. The space $S_2(\Gamma)$ of forms of weight 2 on $\mathcal{H}/\Gamma$ can then be identified with the space of holomorphic differential forms on the compact Riemann surface $\mathcal{H}/\Gamma$.

We now introduce certain subgroups of $\mathbf{SL}_2(\mathbb{R})$ arising from quaternion algebras which will play much the same role in our discussion as the groups $\Gamma_0(N)$ of Chapter 2. Let N be a positive integer.

DEFINITION 4.9. The factorisation $N = N^+N^-$ is called an *admissible factorisation* if

(1) $\gcd(N^+, N^-) = 1$,
(2) the integer N^- is squarefree, and the product of an even number of primes.

A discrete subroup Γ_{N^+,N^-} of $\mathbf{SL}_2(\mathbb{R})$ can be associated to any admissible factorisation of N as follows: let B denote the quaternion algebra ramified precisely at the primes ℓ which divide N^-. (Such an algebra is unique, up to isomorphism, by Proposition 4.1.) Note that B is an *indefinite* quaternion algebra, i.e., it is split at the place ∞.

Choose a maximal order R_0 in B. Such orders are unique up to conjugation by $B^\times$, by Proposition 4.4. Since the algebra B is split at all the primes dividing N^+, and R_0 is a maximal order, one may fix an identification

$$\eta : R_0 \otimes (\mathbb{Z}/N^+\mathbb{Z}) \longrightarrow M_2(\mathbb{Z}/N^+\mathbb{Z}).$$

Let R denote the subring of R_0 consisting of all elements x such that $\eta(x)$ is upper triangular. The subring R is an Eichler order of level N^+ in B. Like the maximal order R_0, the Eichler order R is unique up to conjugation by $B^\times$. After fixing as before an identification ι of $B \otimes \mathbb{R}$ with $M_2(\mathbb{R})$, define

$$\Gamma_{N^+,N^-} = \iota(R_1^\times),$$

where $R_1^\times$ denotes as before the group of elements of reduced norm 1 in R.

We now collect some basic facts about the structure of the space

$$S_2(\Gamma_{N^+,N^-}) =: S_2(N^+,N^-)$$

which are analogous to the basic properties of $S_2(N)$ discussed in Chapter 2:

- The space $S_2(N^+,N^-)$ is naturally a Hilbert space, in which the duality is given by the wedge product of differential one-forms (cup-product).
- It is endowed with a natural action of Hecke operators T_p, indexed by rational primes p, which are *self-adjoint* when p does not divide N. To define T_p in this case, let $\alpha \in R$ be an element of reduced norm p. The double coset $\Gamma\alpha\Gamma$ can be written as a disjoint union of left cosets

$$\Gamma\alpha\Gamma = \bigcup_{i=0}^{p} \alpha_i\Gamma,$$

and T_p is defined by summing the translates of f by the left coset representatives α_i

$$T_p(f(z)dz) := \sum_{i=0}^{p} f(\alpha_i^{-1}z)d(\alpha_i^{-1}z). \tag{4.4}$$

- Because the Hecke operators T_n for $(n,N)=1$ commute and are self-adjoint, the space $S_2(\Gamma_{N^+,N^-})$ is completely diagonalisable under the action of these operators.
- If f is an eigenform for the Hecke operators, its associated L-function can be defined as the product of the following local factors (at least for the primes ℓ which do not divide N):

$$(1 - a_\ell(f)\ell^{-s} + \ell^{1-2s})^{-1}, \quad \text{where } T_\ell f = a_\ell f.$$

REMARK 4.10. We have not said anything about the dimensions of the various eigenspaces, and it should be remarked that here lies a complication of the theory: it is not clear that a simultaneous eigenspace for all the Hecke operators should be one-dimensional, since one lacks the notion of Fourier expansion which in the case of forms on $\Gamma_0(N)$ allows one to recover the eigenform from a knowledge of its associated system of Hecke eigenvalues.

Nonetheless, there is a generalisation of Atkin-Lehner theory in this setting. More precisely, one can define a notion of oldforms in $S_2(\Gamma_{N^+,N^-})$, which are forms arising from forms in $S_2(\Gamma_{d^+,N^-})$ where d^+ is a proper divisor of N^+. The space of newforms is the orthogonal complement of the space of oldforms defined in this way. It is proved in [**Zh01a**], §3.2.1, that the simultaneous eigenspaces in $S_2^{\text{new}}(\Gamma_{N^+,N^-})$ for all the Hecke operators (or even merely for the good Hecke operators) are one-dimensional.

We call a modular form f in such an eigenspace an *eigenform* on Γ_{N^+,N^-}. Since f does not admit a Fourier expansion, it is also unclear by what condition one might normalise f in order to arrive at a notion of *normalised eigenform*. We postpone the discussion of this issue to the next chapter.

4.3. Shimura curves

The compact Riemann surface $\mathcal{H}^*/\Gamma_0(N)$ can be interpreted as the complex points of an algebraic curve $X_0(N)$ *defined over* $\mathbb{Q}$. As was proved by Shimura,

an analogous fact holds for the quotients $\mathcal{H}/\Gamma_{N^+,N^-}$. In fact this Riemann surface admits a moduli interpretation as classifying abelian surfaces over $\mathbb{Q}$ endowed with certain extra stuctures; since this moduli problem makes sense over $\mathbb{Q}$, it gives rise to an algebraic curve X_{N^+,N^-} over $\mathbb{Q}$ whose complex points are identified with $\mathcal{H}/\Gamma_{N^+,N^-}$.

Roughly speaking, the moduli interpretation associates to $\tau \in \mathcal{H}/\Gamma_{N^+,N^-}$ an *abelian surface* with endomorphism ring containing the order R_0, and certain auxiliary level N^+ structure. (For more details on Shimura curves and a precise definition of the moduli problem, see [**BC92**] Chapter 1 of [**Zh01a**], or Chapter 0 of [**Cl03**].)

For example, if $N^- = 1$ so that $B = M_2(\mathbb{Q})$, the maximal order R_0 can be chosen to be $M_2(\mathbb{Z})$. An abelian surface A whose endomorphism ring contains $M_2(\mathbb{Z})$ decomposes as a product of an elliptic curve E with itself:

$$A = E \times E = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} A \times \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} A.$$

The level N structure imposed on A corresponds to the usual level N structure on E, so that in this case one recovers the usual moduli interpretation of $X_0(N)$.

4.4. The Eichler-Shimura construction, revisited

Let f be an eigenform in $S_2(\Gamma_{N^+,N^-})$ having integer Hecke eigenvalues $a_n(f)$. As in the case of modular forms on $\Gamma_0(N)$, one can associate to such an eigenform an elliptic curve over $\mathbb{Q}$:

THEOREM 4.11. *There exists an elliptic curve E over $\mathbb{Q}$ such that $a_n(E) = a_n(f)$, for all integers n such that $(n, N) = 1$.*

SKETCH OF PROOF. The proof proceeds along lines similar to those of theorem 2.10 of Chapter 2. Let $\mathbb{T}$ be the algebra generated by the good Hecke operators. These operators can be realised as algebraic correspondences on the Shimura curves X_{N^+,N^-} and hence give rise to endomorphisms of the Jacobian J_{N^+,N^-} of X_{N^+,N^-} which are defined over $\mathbb{Q}$. The eigenform f determines a homomorphism

$$\varphi_f : \mathbb{T} \longrightarrow \mathbb{Z}, \quad \text{sending } T_n \text{ to } a_n(f).$$

Let I_f denote the kernel of φ_f. The multiplicity one result alluded to in Remark 4.10 implies that the quotient

$$E_f := J_{N^+,N^-}/I_f$$

is an elliptic curve. An analogue of the Eichler-Shimura congruence, this time for the correspondence T_p on $X^2_{N^+,N^-}$, yields the equality of L-functions

$$L(E_f, s) = L(f, s).$$

For more details on this construction see [**Zh01a**], sec. 3.4. □

4.5. The Jacquet-Langlands correspondence

The Eichler-Shimura construction of the previous section, combined with Wiles' theorem that every elliptic curve is modular, leads to the conclusion that for every admissible factorisation N^+N^- of N and for every newform g on Γ_{N^+,N^-} with integer Hecke eigenvalues, there is an associated newform f on $\Gamma_0(N)$ with the same Hecke eigenvalues as those of g at the primes ℓ not dividing N. In fact, more is true, a fact which could be established before Wiles' proof of the Shimura-Taniyama-Weil

conjecture: one does not need to assume the rationality of the Fourier coefficients of f, and the correspondence between newforms goes both ways.

THEOREM 4.12 (Jacquet-Langlands). *Let f be a newform on $\Gamma_0(N)$, and let $N = N^+N^-$ be an admissible factorisation of N. Then there is a newform $g \in S_2(\Gamma_{N^+,N^-})$ with*

$$L(f,s) = L(g,s) \quad \textit{(up to finitely many Euler factors).}$$

The proof of this theorem, which relies on techniques of non-abelian harmonic analysis, is beyond the scope of these notes, and is explained in [**Gel75**] (specifically, in the last chapter) and in [**JL70**].

4.6. The Shimura-Taniyama-Weil conjecture, revisited

The results of Section 4.5 make it possible to rewrite the Shimura-Taniyama-Weil conjecture in terms of modular forms on Γ_{N^+,N^-}.

THEOREM 4.13. *Let $E/\mathbb{Q}$ be an elliptic curve of conductor N, and let $N = N^+N^-$ be an admissible factorisation of N. Then there exists a unique eigenform $f \in S_2(\Gamma_{N^+,N^-})$ such that*

$$T_\ell(f) = a_\ell(E)f, \qquad \textit{for all } \ell \not| N.$$

SKETCH OF PROOF. By Wiles' theorem, there exists a newform f_0 on $\Gamma_0(N)$ attached to E. Theorem 4.12 produces the desired eigenform $f \in S_2(\Gamma_{N^+,N^-})$. □

Theorem 4.13 supplies an essential ingredient in defining the new type of modular parametrisation

$$\Phi_{N^+,N^-} : \mathrm{Div}^0(\mathcal{H}/\Gamma_{N^+,N^-}) \longrightarrow E(\mathbb{C}).$$

To begin, let

$$\Phi^0_{N^+,N^-} : \mathrm{Div}^0(\mathcal{H}) \longrightarrow \mathbb{C}$$

be the map which to a divisor D associates the line integral $\int_D f(z)dz$. The subgroup generated by the elements of the form $\Phi_{N^+,N^-}(D)$, where D is a divisor which becomes trivial in $\mathcal{H}/\Gamma_{N^+,N^-}$, is a lattice Λ_f in $\mathbb{C}$, and $\mathbb{C}/\Lambda_f = E_f(\mathbb{C})$. One thus obtains a map

$$\Phi'_{N^+,N^-} : \mathrm{Div}^0(\mathcal{H}/\Gamma_{N^+,N^-}) \longrightarrow \mathbb{C}/\Lambda_f = E_f(\mathbb{C}).$$

Since E and E_f have the same L-function, they are isogenous over $\mathbb{Q}$. Letting α be an isogeny $E_f \longrightarrow E$ defined over $\mathbb{Q}$, one then sets

$$\Phi_{N^+,N^-} = \alpha\Phi'_{N^+,N^-}.$$

4.7. Complex multiplication for $\mathcal{H}/\Gamma_{N^+,N^-}$

The reader will note that the one-dimensional factors of jacobians of Shimura curves do not yield any new elliptic curves over $\mathbb{Q}$, since these are already all accounted for in the jacobians of the modular curves $X_0(N)$, by Wiles' theorem. However, the larger supply of modular parametrisations

$$\Phi_{N^+,N^-} : \mathrm{Div}^0(\mathcal{H}/\Gamma_{N^+,N^-}) \longrightarrow E(\mathbb{C}),$$

indexed by admissible factorisations of N provide new constructions of algebraic points on E, and in fact examples of Heegner systems that could not be constructed from modular curve parametrisations alone.

Following the lead of Chapter 3 and defining CM points on $\mathcal{H}/\Gamma_{N^+,N^-}$ as arising from $\tau \in \mathcal{H} \cap K$, where K is an imaginary quadratic subfield of $\mathbb{C}$, is clearly inappropriate since the group Γ_{N^+,N^-}, which depends on an identification ι of $B \otimes \mathbb{R}$ with $M_2(\mathbb{R})$, is only well-defined up to conjugation in $\mathbf{SL}_2(\mathbb{R})$, a group whose action does not preserve $\mathcal{H} \cap K$. One resorts to the characterisation of CM points as those whose associated orders are orders in imaginary quadratic fields. More precisely:

DEFINITION 4.14. Given $\tau \in \mathcal{H}/\Gamma_{N^+,N^-}$, the *associated order* of τ is the set

$$\mathcal{O}_\tau := \{\gamma \in R \quad \text{such that } \mathrm{norm}(\gamma) = 0 \text{ and } \iota(\gamma)(\tau) = \tau\} \cup \{0\}.$$

As in the case treated in Chapter 3, the assignment $\gamma \mapsto z_\gamma$ identifies $\mathcal{O}_\tau$ with a discrete subring of $\mathbb{C}$, so that $\mathcal{O}_\tau$ is either $\mathbb{Z}$ or an order in an imaginary quadratic field $K \subset \mathbb{C}$.

DEFINITION 4.15. A point $\tau \in \mathcal{H}/\Gamma_{N^+,N^-}$ is called a CM point if its associated order is isomorphic to an order in an imaginary quadratic field.

As in Chapter 3, given an order $\mathcal{O}$ in an imaginary quadratic field K we write

$$CM(\mathcal{O}) = \{\tau \in \mathcal{H}/\Gamma_{N^+,N^-} \quad \text{such that } \mathcal{O}_\tau = \mathcal{O}\}.$$

The importance of the CM points lies in the fact that the theory of complex multiplication formulated in Chapter 3 in the case of classical modular curves generalises readily to this new setting:

THEOREM 4.16 (Complex multiplication for Shimura curves). *Let $\mathcal{O}$ be an order in an imaginary quadratic field K of discriminant prime to N, and let H/K be the ring class field of K attached to $\mathcal{O}$. Then*

$$\Phi_{N^+,N^-}(\mathrm{Div}^0(CM(\mathcal{O}))) \subset E(H).$$

SKETCH OF PROOF. The proof uses the moduli interpretation of the points on $\mathcal{H}/\Gamma_{N^+,N^-}$. If τ belongs to $CM(\mathcal{O})$, the associated abelian suface A_τ has endomorphisms by the maximal order R_0, as well as by $\mathcal{O}$, and these two actions commute with each other. Hence A_τ has endomorphisms by $R_0 \otimes_{\mathbb{Z}} \mathcal{O}$, and order in $B \otimes K \simeq M_2(K)$. It follows that A_τ is isogenous to a product $A' \times A'$, where A' is an elliptic curve with complex multiplication by $\mathcal{O}$. Hence A_τ is defined over H by the theory of complex multiplication covered in Chapter 3. Further work shows that the level N^+ structure attached to A_τ gives rise to a level N^+ structure on A' which is defined over H as well. □

4.8. Heegner systems

The following lemma reveals that the CM points arising from Shimura curve parametrisations are fundamentally new sets of points that could not be obtained by using modular curve parametrisations alone.

LEMMA 4.17. *Let K be an imaginary quadratic field of discriminant prime to N and let $\mathcal{O}$ be an order in K of conductor prime to N. Then $CM(\mathcal{O}) \neq \emptyset$ if and only if the following two conditions are satisfied:*

1. *All the primes ℓ dividing N^- are inert in K;*
2. *All the primes ℓ dividing N^+ are split in K.*

PROOF. Since K is a quadratic subfield of the quaternion algebra B which is ramified at N^-, it follows that all the primes dividing N^- are inert in K. The fact that all primes dividing N^+ are split in K is proved exactly as in the proof of Proposition 3.8 of Chapter 3. □

Lemma 4.17 leads to the proof of the following theorem.

THEOREM 4.18. *Let E be a semistable elliptic curve of conductor N, and let K be an imaginary quadratic field of discriminant prime to N. If* $\mathrm{sign}(E,K) = -1$, *then there is a non-trivial Heegner system $\{P_n\}$ attached to (E,K).*

SKETCH OF PROOF. The field K determines a factorisation $N = N^+N^-$ of N by letting N^+ be the product of the primes which are split in K, while N^- is the product of the primes which are inert in K. Since $\mathrm{sign}(E,K) = -1$, the set $S_{E,K}$ has odd cardinality. On the other hand,

$$\begin{aligned} S_{E,K} &= \{\lambda|\ell|N^+ \text{ such that } E/\mathbb{Q}_\ell \text{ has split multiplicative reduction at } \ell\} \\ &\quad \cup \quad \{\ell|N^-\} \quad \cup \quad \{\infty\}. \end{aligned}$$

The first set in the union has even cardinality, hence it follows that N^- is divisible by an even number of primes as well, so that N^+N^- is an admissible factorisation of N. For each integer n which is prime to N, one then knows by Lemma 4.17 that $CM(\mathcal{O}_n)$ is non-empty. One may choose divisors $D_n \in \mathrm{Div}^0(CM(\mathcal{O}_n))$ in such a way that $P_n := \Phi_{N^+,N^-}(D_n)$ forms a Heegner system. An argument analogous to the proof of Theorem 3.13 of Chapter 3, based on the density of the CM points in $\mathcal{H}/\Gamma_{N^+,N^-}$ ensures that this Heegner system is non-trivial. □

4.9. The Gross-Zagier formula

An analogue of the Gross-Zagier formula (Theorem 3.20) for Heegner points which arise from Shimura curve parametrisations was anticipated by Gross and Zagier in [**Gr84**] and has been recently proved by Zhang [**Zh01a**].

THEOREM 4.19 (Zhang). *If $\{P_n\}$ is the Heegner system attached to (E,K) as above, and if $P_K := Trace_{H_1/K}(P_1)$, then*

$$\langle P_K, P_K \rangle \doteq L'(E/K, 1).$$

(The symbol $\doteq$ is given the same meaning here as in the statement of Theorem 3.20.)

We close this chapter by raising two questions which arise naturally from our discussion of Shimura curves:

(1) How does one compute numerically the parametrisation Φ_{N^+,N^-} when $N^- \neq 1$? The Fourier expansion of the modular form f attached to E in a neighbourhood of $i\infty$ when $N^- = 1$ does not generalise in any obvious way to the setting of Shimura curves which are not equipped with cusps.
(2) The second question is the primary motivation for Chapters 6,7, and 8: What construction plays the role of modular and Shimura curve parametrisations, and of the CM points on these curves, when the field K is real quadratic? In that setting, is it possible to construct the Heegner systems whose existence is predicted by Conjecture 3.16 when $\mathrm{sign}(E,K) = -1$?

A partial answer to question 1 can be given by exploiting a structure on Shimura curves which has no counterpart for classical modular curves: the p-adic uniformisation of these curves by certain discrete arithmetic subgroups of $\mathbf{SL}_2(\mathbb{Q}_p)$, for p a prime dividing N^-. This new structure in some ways *compensates* for the absence of cusps and Fourier expansions, in allowing an explicit numerical description of modular forms in $S_2(\Gamma_{N^+,N^-})$.

Question 2 lies deeper. A partial conjectural answer to it is given in Chapter 9, relying on modular symbols and on the p-adic analytic techniques introduced in the next chapter.

References

The behaviour of the Mordell-Weil groups $E(H)$ when E is an elliptic curve of prime conductor p and H is a ring class field of an imaginary quadratic field K in which p is inert, so that $\mathrm{sign}(E,K)=1$, is studied in [**BD97**] where it is proved that $E(H)$ is finite if $L(E,H,1)\neq 0$, for any ring class field H of conductor prime to p. More precisely, if $G=\mathrm{Gal}(H/K)$ and $\chi:\mathbb{Z}[G]\longrightarrow\mathbb{C}$ is an algebra homomorphism, the "χ-part" $E(H)^\chi := E(H)\otimes_\chi \mathbb{C}$ is trivial when $L(E/K,\chi,1)\neq 0$. When $\mathrm{sign}(E,K)=1$, the non-vanishing of the factors $L(E/K,\chi,1)$ of $L(E,H,1)$ is expected to occur "most of the time". For example, work of Cornut and Vatsal [**Cor02**], [**Va02**] shows that $L(E/K,\chi,1)\neq 0$ for almost all ring class characters whose conductor is a power of a fixed prime ℓ. It follows that there is no Heegner system attached to (E,K) in this situation.

The book of Vigneras [**Vi80**] is a good reference for the arithmetic of quaternion algebras and its connection with modular forms. The arithmetic theory of modular forms on quaternion algebras and of Shimura curves has been developed by Shimura [**Sh67**] building on earlier work of Eichler.

In addition to the canonical models of X_{N^+,N^-} over $\mathbb{Q}$ provided by Shimura's theory, one also has at one's disposal integral models which have good reduction outside N. (Cf. [**Car86**], [**Dr76**].)

A useful reference for the topic of Heegner points on Shimura curves are the articles [**Zh01a**] and [**Zh01b**] which prove the Gross-Zagier formula in the context of Shimura curves in a rather general setting and contain helpful background on modular forms attached to quaternion algebras.

Exercises

(1) Let E be an elliptic curve of conductor N and let K be an imaginary quadratic field in which all the primes dividing N are inert. Let $\Phi_N:\mathcal{H}/\Gamma_0(N)\longrightarrow E(\mathbb{C})$ be the classical modular parametrisation attached to E.
   (a) Suppose that $N=p$ is prime. Show that that if τ belongs to $\mathcal{H}\cap K$, then $P_\tau:=\Phi_N(\tau)$ belongs to $E(H_n)$ for some n of the form $p^t n'$ with $t\geq 1$ and $(p,n')=1$. Furthermore, show that

$$\mathrm{Trace}_{H_n/H_{n'}}(P_\tau)=0. \tag{4.5}$$

   (b) Suppose that $N=pq$ is the product of two distinct primes. Show that the points of the form $\Phi_N(\tau)$ are defined over ring class fields of conductor $n=p^tq^sn'$ with $t,s\geq 1$, and that equation (4.5) continues to hold, even though $\mathrm{sign}(E,K)=-1$ in this case. This justifies working with the

Shimura curve parametrisation $\Phi_{1,pq}$ to produce the non-trivial Heegner system whose existence is predicted in this case.

(2) Let B be a quaternion algebra over a field F.
   (a) If $\alpha \in B \setminus F$, show that the subalgebra $K = F(\alpha)$ generated by α over F is a commutative semisimple algebra of rank 2, and that it is a field if B is a division algebra. Let $x \mapsto \bar{x}$ denote the involution of K/F.
   (b) Fix the quadratic subalgebra $K \subset B$. Show that there is an element $\beta \in B^\times$ satisfying $\beta\lambda = \bar{\lambda}\beta$ for all $\lambda \in K$. (Hint: Study the K-linear action of K by right multiplication on B viewed as a K-vector space under left multiplication.) Show that the element β is uniquely determined by K up to multiplication by elements of $K^\times$.
   (c) Show that $\gamma = \beta^2$ belongs to F, and that it is uniquely determined up to multiplication by norms of non-zero elements in K.
   (d) Conclude that any quaternion algebra over F is isomorphic to an algebra of the form $B_{K,\gamma} = \{a + b\beta | a, b \in K\}$, where K is a quadratic semisimple algebra over F and $\gamma \in F$, with multiplication given by the rule
$$(a + b\beta)(a' + b'\beta) = (aa' + b\bar{b}'\gamma) + (ab' + b\bar{a}')\beta.$$
   (e) Show that the only quaternion algebras over $\mathbb{R}$ are the split algebra $M_2(\mathbb{R})$ and the algebra of Hamilton quaternions.

(3) Let B be a quaternion algebra over F, and let $b \in B$. Let K be a subfield of B quadratic over F and containing b. Prove that the norm and trace of b from K to F are equal to their reduced norm and trace from B.

(4) Show that a quaternion algebra becomes split over any quadratic subfield.

(5) Let B be a quaternion algebra over a global field. Show that the *level* of an Eichler Z-order $R = R_1 \cap R_2$, defined as the Z-module index of R in either R_1 or R_2, is independent of the expression of R as the intersection of two maximal orders R_1 and R_2. (Hint: prove this first for orders in a matrix algebra over a local field.)