Rational points on modular elliptic curves

Henri Darmon

Department of Mathematics, McGill University, Montreal, Quebec, Canada, H3A-2K6

Current address: Department of Mathematics, McGill University, Montreal, Quebec, Canada H3A-2K6

E-mail address: darmon@math.mcgill.ca

1991 *Mathematics Subject Classification.* Primary 11-02; Secondary 11F03 11F06 11F11 11F41 11F67 11F75 11F85 11G05 11G15 11G40

Key words and phrases. Elliptic curves, modular forms, Heegner points, Shimura curves, rigid analysis, p-adic uniformisation, Hilbert modular forms, Stark-Heegner points, Kolyvagin's theorem.

Some of the ideas discussed in this monograph have their roots in the author's collaboration with Massimo Bertolini over the years. Exchanges with Samit Dasgupta, Peter Green, Adrian Iovita, and Adam Logan have also helped in forming and solidifying key insights. The author thanks Drs. Peter Hilton and Heath Martin of the University of Central Florida for the marvelous job they did in running the NSF-CBMS conference on which this monograph is based, as well as the participants for their many stimulating comments and suggestions. A rough version of this text formed the basis for a graduate seminar at McGill University in the 2002-2003 academic year. The participants of this seminar – Samit Dasgupta, Hugo Chapdelaine, Antoine Gournay, and Matt Greenberg – pointed out a number of mistakes and their suggestions have led to improvements in the manuscript. (Needless to say, the imperfections and inaccuracies which remain are to be blamed on the author alone!) Finally, it is a great pleasure to acknowledge NSERC and CICMA whose material support has greatly facilitated the writing of this monograph.

Abstract. Based on an NSF-CBMS lecture series given by the author at the University of Central Florida in Orlando from August 8 to 12, 2001, this monograph surveys some recent developments in the arithmetic of modular elliptic curves, with special emphasis on the Birch and Swinnerton-Dyer conjecture, the construction of rational points on modular elliptic curves, and the crucial role played by modularity in shedding light on these questions.

Pour Galia et Maia

Contents

Preface

This monograph is based on an NSF-CBMS lecture series given by the author at the University of Central Florida in Orlando from August 8 to 12, 2001.

The goal of this lecture series was to survey some recent developments in the arithmetic of modular elliptic curves, with special emphasis on

(1) the Birch and Swinnerton-Dyer conjecture;
(2) the construction of rational points on modular elliptic curves;
(3) the crucial role played by modularity in shedding light on these two closely related issues.

The text is divided into three parts of roughly equal length.

The first consists of chapters 1-3 and chapter 10. The first three chapters introduce the background and prerequisites for what follows: elliptic curves, modular forms and the Shimura-Taniyama-Weil conjecture, complex multiplication and the fundamental *Heegner point construction* whose study and generalisation is the main theme of the monograph. The notion of "Heegner system" which is spelled out in chapter 3 is used in chapter 10 to prove Kolyvagin's theorem relating Heegner points to the arithmetic of elliptic curves, giving strong evidence for the Birch and Swinnerton-Dyer conjecture for elliptic curves of analytic rank at most one. While more advanced than chapters 1-3, chapter 10 is independent of the material in chapters 4-9 and could be read immediately after chapter 3.

Chapters 4-6 introduce variants of modular parametrisations in which modular curves are replaced by Shimura curves attached to certain indefinite quaternion algebras. A study of these parametrisations reveals an important new structure: the rigid analytic uniformisation of Shimura curves discovered by Cerednik and Drinfeld, giving rise to p-adic uniformisations of modular elliptic curves by discrete arithmetic subgroups of $\mathbf{SL}_2(\mathbb{Q}_p)$ arising from definite quaternion algebras.

The main new contributions of this monograph are contained in chapters 7-9. These chapters give an overview of the author's attempts to extend the theory of Heegner points and complex multiplication to certain situations where the base field is not a CM field. The notions of rigid analysis developed in chapters 5 and 6 play a key role in suggesting a p-adic variant of the theories of chapters 7 and 8. This leads, in chapter 9, to a conjectural construction of points on a modular elliptic curve over $\mathbb{Q}$ defined over ring class fields of a *real* quadratic field, which are expected to behave much like classical Heegner points attached to an imaginary quadratic field.

The reader is cautioned that many proofs give only the main ideas; details have often been left out or relegated to exercises, retaining (for better or for worse) the flavour of the original lecture series. Of necessity, a number of important topics had to be omitted or inadequately touched upon, for lack of time. The material covered here would be suitable for a 10 to 15-hour mini-course, or, with extra background material, for a one-semester or even a year-long seminar aimed at graduate students.

A selection of exercises is given at the end of each section. Many of these consist in working out the details of arguments sketched in the text. It is hoped that readers encountering this material for the first time will find the exercises helpful, while more sophisticated readers may elect to skip them without loss of continuity.

Henri Darmon

Montreal 2003

CHAPTER 1

Elliptic curves

1.1. Elliptic curves

DEFINITION 1.1. An elliptic curve over a field F is a complete algebraic group over F of dimension 1.

Equivalently, an elliptic curve is a smooth projective curve of genus one over F equipped with a distinguished F-rational point, the identity element for the algebraic group law. It is a consequence of the Riemann-Roch theorem ([**Si86**], chap. III) that when F is of characteristic different from 2 and 3 such a curve can be described by an affine equation of the form

$$E : y^2 = x^3 + ax + b, \quad \text{with } a, b \in F, \quad \Delta := -2^4(4a^3 + 27b^2) \neq 0, \tag{1.1}$$

in which the distinguished F-rational point is taken to be the unique point at infinity in the homogeneous equation for the corresponding projective curve. Over a field of arbitrary characteristic, an elliptic curve can still be described as a plane cubic curve, given by the somewhat more complicated equation sometimes referred to as the *generalised Weierstrass normal form*

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \quad \text{with } \Delta \neq 0. \tag{1.2}$$

In fact, elliptic curves are sometimes *defined* as plane cubic curves given by an equation of the form (1.1) or (1.2), the addition law on E being described geometrically via the well-known chord-and-tangent law. Such an approach has the virtue of concreteness and underscores the possibility of doing explicit calculations with elliptic curves (by computer, or even by hand) which is one of the charms of the subject. On the other hand, definition 1.1 is more conceptual and explains why elliptic curves should be singled out for special attention: they are the only projective curves that can be endowed with an algebraic group law - a structure which both facilitates and enriches their diophantine study.

For two elliptic curves over F to be isomorphic over $\bar{F}$, it is necessary and sufficient that their so-called *j-invariants*, defined in terms of the coefficients of equation (1.1) by the formula

$$j = -\frac{2^{12}3^3a^3}{\Delta} \tag{1.3}$$

be equal.

The structure of the group $E(F)$ of solutions to (1.1) or (1.2) depends of course on the nature of the field F, for example:

1. When F is a finite field, the group $E(F)$ is a finite abelian group. The study of $E(F)$ is at the origin of many of the practical applications of elliptic curves to cryptography and coding theory.

2. If F is the real or complex numbers, then $E(F)$ inherits from F the structure of a compact abelian group. For example, when $F = \mathbb{R}$, the group $E(F)$ is abstractly isomorphic either to a circle group, or the product of a circle and $\mathbb{Z}/2\mathbb{Z}$. If $F = \mathbb{C}$ is the field of complex numbers, then $E(\mathbb{C})$ is topologically isomorphic to a torus. As a complex analytic manifold it is isomorphic to the quotient of $\mathbb{C}$ by a lattice $\Lambda \subset \mathbb{C}$ generated by the periods of a holomorphic differential ω against the integral homology of $E(\mathbb{C})$. To make the isomorphism $\mathbb{C}/\Lambda \longrightarrow E(\mathbb{C})$ explicit, let $\wp_\Lambda(z)$ be the Weierstrass $\wp$-function attached to Λ, defined by

$$\wp_\Lambda(z) = \frac{1}{z^2} + \sum_{\lambda \in \Lambda - \{0\}} \left(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2} \right).$$

The Λ-periodic functions $x = \wp_\Lambda(z)$ and $y = \wp'_\Lambda(z)$ satisfy the algebraic relation

$$y^2 = 4x^3 - g_2 x - g_3, \tag{1.4}$$

where

$$g_2 = 60 \sum_{\lambda \in \Lambda - \{0\}} \frac{1}{\lambda^4}, \quad g_3 = 140 \sum_{\lambda \in \Lambda - \{0\}} \frac{1}{\lambda^6}, \tag{1.5}$$

and the map

$$\Phi_w(z) = (\wp_\Lambda(z), \wp'_\Lambda(z)) \tag{1.6}$$

gives an isomorphism (of groups as well as complex analytic varieties) between $\mathbb{C}/\Lambda$ and the elliptic curve with equation (1.4), which is isomorphic to E over $\mathbb{C}$. For more details, see [**Si86**], chap. VI.

The isomorphism between $E(\mathbb{C})$ and $\mathbb{C}/\Lambda$ makes it transparent that the group E_n of points on E of order n, with coordinates in $\mathbb{C}$ or in any algebraically closed field of characteristic zero, is isomorphic to $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ as an abstract group.

3. If F is a p-adic field ($\mathbb{Q}_p$, say, or a finite extension of $\mathbb{Q}_p$) then $E(F)$ is a compact p-adic Lie group, hence an extension of a finite group by a pro-p group $E_1(F)$ (cf. [**Si86**], chaps. IV and VII).

One may assume, after a change of variables, that E is given by an equation of the form (1.2) in which the coefficients a_i belong to the ring of integers $\mathcal{O}_F$ of F, and for which the associated discriminant $\Delta_{min} = \Delta$ has minimal valuation in $\mathcal{O}_F$. If Δ belongs to $\mathcal{O}_F^\times$, then the equation obtained by reducing (1.2) modulo a uniformiser $\pi \in \mathcal{O}_F$ defines an elliptic curve over the finite field $k = \mathcal{O}_F/(\pi)$. In this case one says that E/F has "good reduction".

An important role is played in this monograph by elliptic curves having a special type of bad reduction, referred to as *multiplicative reduction*. This is the case where $\Delta_{min} \in \mathcal{O}_F$ is not a unit and where the equation obtained by reducing (1.2) modulo π has an ordinary double point as its only singularity. In that case $\mathrm{ord}_\pi(j) < 0$, and there is a $q \in F^\times$ which can be obtained by formally inverting the power series expressing j in terms of q

$$j = \frac{1}{q} + 744 + 196884q + \cdots$$

to express q as a power series in $1/j$ with integral (and hence p-adically bounded) coefficients. The curve E is isomorphic over $\bar{F}$ (more precisely, over a quadratic unramified extension of F) to the curve E_q given by the equation

$$E_q : y^2 + xy = x^3 + a_4(q)x + a_6(q), \tag{1.7}$$

where

$$a_4(q) = -s_3(q), \quad a_6(q) = \frac{-5s_3(q) + 7s_5(q)}{12}, \qquad \text{with } s_k(q) = \sum_{n=1}^{\infty} \frac{n^k q^n}{1-q^n}. \tag{1.8}$$

The p-adic analytic isomorphism $\Phi_{\mathrm{Tate}} : \bar{F}^\times \longrightarrow E_q(\bar{F})$ is obtained by setting

$$\Phi_{\mathrm{Tate}}(u) = \left(\sum_{n\in\mathbb{Z}} \frac{q^n u}{(1-q^n u)^2} - 2s_1(q), \sum_{n\in\mathbb{Z}} \frac{(q^n u)^2}{(1-q^n u)^3} + s_1(q) \right). \tag{1.9}$$

(For more details, see [**Si94**], ch. V.) This p-adic uniformisation theory for E yields a description of $E(\bar{F})$ in the spirit of the Weierstrass theory of equation (1.6). It plays an important role in the constructions of chapters 6 and 9.

4. The study of elliptic curves over the finite, complex and p-adic fields, while of interest in its own right, is subordinate in this monograph to the case where F is a *number field* – the field of rational numbers or a finite extension of it. The key result and the starting point for the theory of elliptic curves over number fields is the *Mordell-Weil theorem.* This result was first established, in certain special cases, by Fermat himself using his method of descent, and its proof is recalled in the next section.

1.2. The Mordell-Weil theorem

Let E be an elliptic curve defined over a number field F.

THEOREM 1.2 (Mordell-Weil). *The Mordell-Weil group $E(F)$ is finitely generated, i.e.,*

$$E(F) \simeq \mathbb{Z}^r \oplus T,$$

where $r \geq 0$ and T is a finite group.

Since (the modern formulation of) the proof of the Mordell-Weil theorem plays an important role in the questions studied in this monograph, particularly in chapter 10, it is worthwhile to recall here the main ideas which are behind it.

PROOF OF THEOREM 1.2. The proof is composed of two ingredients:

1. The existence of a *height function* $h : E(F) \longrightarrow \mathbb{R}$ satisfying suitable properties:

THEOREM 1.3. *There exists a function*

$$h : E(F) \longrightarrow \mathbb{R}$$

satisfying

(1) *For all points Q in $E(F)$, there is a constant C_Q depending only on Q, and an absolute constant C depending only on E, such that*

$$h(P+Q) \leq 2h(P) + C_Q, \quad h(mP) \geq m^2 h(P) + C,$$

for all $P \in E(F)$.

(2) *For all $B > 0$,*

$$\{P \text{ such that } h(P) < B\} \text{ is finite.}$$

2. The *weak Mordell-Weil theorem*:

THEOREM 1.4. *For any integer $n \geq 1$, the group $E(F)/nE(F)$ is finite.*

These two ingredients are combined in the following *descent lemma* of Fermat of which the Mordell-Weil theorem is a direct consequence:

LEMMA 1.5 (Fermat). *Let G be an abelian group equipped with a height function satisfying the properties of theorem 1.3, and assume that the quotient G/nG is finite for some $n > 1$. Then G is finitely generated.*

It is not our intention to focus any further on heights (which are discussed at length in [**Si86**] for example), or on the (elementary) descent lemma, whose proof is relegated to the exercises. The weak Mordell-Weil theorem is the most interesting ingredient from the point of view of a study of the Birch and Swinnerton-Dyer conjecture, since it is the source of the *non-effectivity* in the proof of the Mordell-Weil theorem.

The proof of theorem 1.4 begins with the observation that this theorem is trivially true over an algebraic closure $\bar{F}$ of F, since the multiplication by n map is surjective on $E(\bar{F})$. Hence the sequence

$$0 \longrightarrow E_n \longrightarrow E(\bar{F}) \xrightarrow{n} E(\bar{F}) \longrightarrow 0. \tag{1.10}$$

of modules equipped with their natural continuous action of $G_F := \mathrm{Gal}(\bar{F}/F)$ is exact. Following the usual conventions of Galois cohomology, denote by

$$H^i(F, M) := H^i(G_F, M)$$

the group of continuous i-cocycles modulo the group of continuous i-coboundaries with values in the G_F-module M.(For details on these definitions, see [**CF67**], chapter IV, or [**Si86**], appendix B). Taking the Galois cohomology of the exact sequence (1.10) gives rise to the long exact cohomology sequence

$$0 \longrightarrow E_n(F) \longrightarrow E(F) \xrightarrow{n} E(F) \xrightarrow{\delta} H^1(F, E_n) \longrightarrow H^1(F, E) \xrightarrow{n} H^1(F, E)$$

from which can be extracted the so-called *descent exact sequence*

$$0 \longrightarrow E(F)/nE(F) \xrightarrow{\delta} H^1(F, E_n) \longrightarrow H^1(F, E)_n \longrightarrow 0. \tag{1.11}$$

The connecting homomorphism δ embeds the group $E(F)/nE(F)$ into an object of Galois-theoretic nature, since $H^1(G_F, E_n)$ depends only on the structure of G_F and of the G_F-module E_n, not on the elliptic curve E itself. For example, if all the n-division points of $E(\bar{F})$ are defined over F so that G_F acts trivially on E_n, then elements in

$$H^1(F, E_n) = \hom(G_F, E_n)$$

are indexed by pairs (L, φ) where L is a finite Galois extension of F and φ is an identification of $\mathrm{Gal}(L/F)$ with a subgroup of E_n. If $H^1(F, E_n)$ were a finite group, the mere existence of the exact sequence (1.11) would be enough to conclude the proof of theorem 1.4. However, this is *never* the case when F is a number field and $n > 1$. It is therefore necessary to exploit local information to pin down the image of δ in $H^1(F, E_n)$ with greater accuracy. More precisely, for any place v of F (archimedean or not) the embedding of F into the completion F_v at the place v, extended to an embedding of $\bar{F}$ into $\bar{F}_v$, induces an inclusion $G_{F_v} \subset G_F$. The exact sequence (1.11) has a local counterpart with F replaced by F_v which fits into the commutative diagram

(1.12)
$$\begin{array}{ccccccccc} 0 & \longrightarrow & E(F)/nE(F) & \xrightarrow{\delta} & H^1(F,E_n) & \longrightarrow & H^1(F,E)_n & \longrightarrow & 0 \\ & & \downarrow & & res_v \downarrow & \searrow \partial_v & \downarrow res_v & & \\ 0 & \longrightarrow & E(F_v)/nE(F_v) & \xrightarrow{\delta} & H^1(F_v,E_n) & \longrightarrow & H^1(F_v,E)_n & \longrightarrow & 0 \end{array}$$

in which the vertical arrows correspond to the natural restriction maps. Since $\delta(E(F)/nE(F))$ is contained in the kernel of ∂_v for all v, it is contained in the so-called *n-Selmer group* of E over F defined as follows:

DEFINITION 1.6. Let E be an elliptic curve over a number field F.

(1) The n-Selmer group of E over F, denoted $\mathrm{Sel}_n(E/F)$, is the set of classes $c \in H^1(F,E_n)$ satisfying $\partial_v(c) = 0$, for all places v of F.
(2) The Shafarevich-Tate group of E/F, denoted $\text{Ш}(E/F)$, is the set of classes $c \in H^1(F,E)$ satisfying $\mathrm{res}_v(c) = 0$, for all places v of F.

The exact sequence (1.11) can now be replaced by the exact sequence involving the n-Selmer group and the n-torsion in the Shafarevich-Tate group:

$$0 \longrightarrow E(F)/nE(F) \xrightarrow{\delta} \mathrm{Sel}_n(E/F) \longrightarrow \text{Ш}(E/F)_n \longrightarrow 0. \tag{1.13}$$

The weak Mordell-Weil theorem is then a consequence of the following general finiteness theorem for the Selmer group:

PROPOSITION 1.7. *The Selmer group* $\mathrm{Sel}_n(E/F)$ *is finite.*

SKETCH OF PROOF OF PROPOSITION 1.7. The proof can itself be divided into two stages: a local study, in which it is shown that $\mathrm{Sel}_n(E/F)$ is contained in the group $H^1_{n\Delta}(F,E_n)$ consisting of cohomology classes $c \in H^1(F,E_n)$ whose restriction to the inertia group I_v at v is trivial, for all places v not dividing $n\Delta$. It is only at this stage of the argument that certain facts about the geometry and arithmetic of elliptic curves (albeit, over local fields) are needed. (Cf. exercise 7.)

A global study is then needed to show that $H^1_{n\Delta}(F,E_n)$ is finite. The key ingredient in this finiteness result (cf. exercise 8) is the Hermite-Minkowski theorem asserting that there are only finitely many extensions of a given number field with bounded degree and ramification. □

To recapitulate, the proof of the Mordell-Weil theorem sketched above (and, in particular, the proof of the weak Mordell-Weil theorem) has led to the introduction of two fundamental invariants, the n-Selmer group of E/F and the Shafarevich-Tate group of E/F, fitting into the fundamental exact sequence (1.13). The weak Mordell-Weil theorem follows from the finiteness of $\mathrm{Sel}_n(E/F)$ whose proof in turn relies on the Hermite-Minkowski theorem, which is itself one of the key general finiteness results of algebraic number theory. □

Since $\mathrm{Sel}_n(E/F)$ is *effectively calculable*, the following question emerges naturally from the proof of theorem 1.2.

QUESTION 1.8. *How good an approximation to* $E(F)/nE(F)$ *is* $\mathrm{Sel}_n(E/F)$, *i.e., how large can* $\text{Ш}(E/F)_n$ *be?*

The following basic conjecture implies that the group $E(F)/nE(F)$ and the n-Selmer group $\mathrm{Sel}_n(E/F)$ only deviate by an amount which is *bounded independently of* n and in fact that these two groups are equal for all but finitely many primes n:

CONJECTURE 1.9 (Shafarevich-Tate). *The group* $Ш(E/F)$ *is finite.*

One is now led to formulate the diophantine problem which this monograph is motivated by and largely devoted to:

QUESTION 1.10. *Given an elliptic curve* E *over a number field* F*, is there an algorithm to*

(1) *Test if* $E(F)$ *is infinite?*
(2) *Compute the rank* r *of the Mordell-Weil group* $E(F)$*?*
(3) *Produce a set* $P_1, \ldots, P_r$ *of generators for* $E(F)/E(F)_{\mathrm{tor}}$*?*

These questions are ostensibly arranged in order of increasing difficulty, but some of the key mysteries are already present in (1).

1.3. The Birch and Swinnerton-Dyer conjecture

The main theoretical insights concerning question 1.10 are derived from the fundamental Birch and Swinnerton-Dyer conjecture.

Much is gained in simplicity, and comparatively little is lost in generality, by restricting to the case where $F = \mathbb{Q}$, the field of rational numbers. In that case the elliptic curve E can be described by a so-called *minimal Weierstrass equation*

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{1.14}$$

in which the coefficients a_i are integers and in which Δ is minimal over all equations of this sort describing E. The invariant Δ attached to this equation is called the *minimal discriminant* of E. The invariant holomorphic differential form

$$\omega_E = \frac{dx}{2y + a_1x + a_3} \tag{1.15}$$

is called the *Néron differential* attached to E.

If p does not divide Δ, then the curve over $\mathbb{F}_p$ obtained by reducing equation (1.14) is an elliptic curve whose isomorphism type does not depend on the choice of a minimal Weierstrass equation. In that case one says that E has *good reduction* at p and denotes by N_p the cardinality of the group $E(\mathbb{F}_p)$ of points on E with coordinates in $\mathbb{F}_p$. The function $p \mapsto N_p$ is a subtle arithmetic function – anticipating somewhat on the discussion of the next chapter, the key *modularity property* yields a device for gaining some control on the behaviour of this function. For now, the main (relatively) elementary information of which one disposes concerning the behaviour of N_p is that it can be written in the form

$$N_p = p + 1 - a_p,$$

where the "error term" a_p satisfies Hasse's inequality (cf. [**Si86**], chapter V):

$$|a_p| \leq 2\sqrt{p}.$$

The basic insight behind the Birch and Swinnerton-Dyer conjecture is that the rank of $E(\mathbb{Q})$ ought to be reflected in the finer properties of the asymptotic behaviour of the quantities N_p as $p \longrightarrow \infty$, and that a large rank, giving rise to an abundant supply of rational points in $E(\mathbb{Q})$, would tend to make the numbers N_p rather larger than $(p+1)$ on average. On the basis of numerical experiments Birch and Swinnerton-Dyer were led to the following conjecture:

CONJECTURE 1.11. *There exists a constant C_E depending only on E such that*

$$\prod_{p<X} \frac{N_p}{p} \simeq C_E(\log X)^r,$$

where r is the rank of $E(\mathbb{Q})$. (Here the symbol $\simeq$ means that the ratio of the expressions apparing on the left and right tends to 1 as $X \longrightarrow \infty$.)

Conjecture 1.11 can be viewed as an example of a "local-global principle" since it asserts that the local invariants N_p "know about" the rank of $E(\mathbb{Q})$, a global invariant.

One of the difficulties of conjecture 1.11 is that the product appearing on the left hand side, involving the arithmetically defined quantities N_p, is hard to come to terms with analytically. A better conceptual understanding of the Birch and Swinnerton-Dyer conjecture can be gained by recasting it in terms of the L-function of $E/\mathbb{Q}$.

1.4. L-functions

To begin with it is useful to extend the definition of the coefficients a_p, which in the previous section were defined only for the primes p not dividing Δ, to a quantity a_n indexed by arbitrary positive integers n.

The definition of a_p is extended to the primes p of bad reduction for E according to a recipe which depends on the type of reduction of E modulo p:

Additive reduction: This occurs if E has a nodal singularity (cusp) so that its group of non-singular points is isomorphic to the additive group G_a. In that case set $a_p := 0$.

Split multiplicative reduction: This occurs if E has an ordinary double point as its only singularity, with tangent lines having rational slopes over $\mathbb{F}_p$. In that case, set $a_p := 1$.

Non-split multiplicative reduction: This occurs if E has an ordinary double point as its only singularity over $\mathbb{F}_p$ with tangent lines having slopes defined over the quadratic extension of $\mathbb{F}_p$ but not $\mathbb{F}_p$ itself. In that case, set $a_p := -1$.

Closely related to the discriminant, but of even greater importance in describing the behaviour of the L-series of E, is the so-called *arithmetic conductor* N of E, which is (almost) characterised by the following three properties:

$$\text{ord}_p(N) = 0 \text{ if and only if } E \text{ has good reduction at } p. \tag{1.16}$$

In particular, N has the same set of prime divisors as the minimal discriminant Δ.

$$\text{ord}_p(N) = 1 \text{ if and only if } E \text{ has multiplicative reduction at } p. \tag{1.17}$$

$$\text{ord}_p(N) = 2 \text{ otherwise, when } p > 3. \tag{1.18}$$

In the remaining case not covered by (1.16), (1.17) and (1.18) where $p = 2$ or 3 and E has additive reduction at p, a more involved recipe for calculating $\text{ord}_p(N)$ is described by an algorithm of Tate [**Ta72**].

The L-function of E is defined as an infinite Euler product

$$L(E,s) = \prod_{p\nmid N}(1 - a_p p^{-s} + p^{1-2s})^{-1} \prod_{p|N}(1 - a_p p^{-s})^{-1} =: \sum a_n n^{-s}, \tag{1.19}$$

in which the expression of $L(E,s)$ as a Dirichlet series supplies the *definition* of the coefficient a_n when n is not prime.

Note that evaluating the Euler product *formally* at $s=1$ gives

$$L(E,1) \text{``}=\text{''} \prod_p \frac{p}{N_p},$$

where N_p is the cardinality of the group of non-singular points in $E(\mathbb{F}_p)$. This equality is just formal, since by exercise 10 the Euler product defining $L(E,s)$ only converges in the right half-plane $\mathrm{Re}(s) > 3/2$.

It is believed that the behaviour of $L(E,s)$ at $s=1$ – assuming one can make sense of it, via the process of analytic continuation – should capture the asymptotics of the product $\prod_{p<X} \frac{N_p}{p}$ appearing in the crude version of the Birch and Swinnerton-Dyer conjecture. This belief is made precise in the following more widely used form of the conjecture:

CONJECTURE 1.12 (Birch and Swinnerton-Dyer). *The L-function $L(E,s)$ extends to an entire function on $\mathbb{C}$ and*

BSD1: *$L(E,1) \neq 0$ if and only if $\#E(\mathbb{Q}) < \infty$.*

BSD2: *The rank r of $E(\mathbb{Q})$ is equal to the order of vanishing of $L(E,s)$ at $s=1$.*

BSD3: *Let R_E be the* regulator *of E, defined by*

$$R_E = \det(\langle P_i, P_j \rangle)_{1\leq i,j\leq r} \#E(\mathbb{Q})_{\mathrm{tors}}^{-2},$$

where $P_1, \ldots, P_r$ is a set of independent generators for $E(\mathbb{Q})/E(\mathbb{Q})_{\mathrm{tors}}$, and $\langle\ ,\ \rangle$ is the canonical Néron-Tate height *attached to $E/\mathbb{Q}$, defined in* [**Si86**], *chapter VIII, §9. Let c_p denote the local terms defined in* [**Si86**], *Chapter VII §6, which depend only on the behaviour of E over $\mathbb{Q}_p$. Then*

$$\lim_{s\to 1} \frac{L(E,s)}{(s-1)^r} = \#Ш(E/\mathbb{Q}) R_E (\prod_{p|N} c_p) c_\infty.$$

Conjecture 1.12 proposes three versions of the Birch and Swinnerton-Dyer conjecture in order of increasing strength. Statement BSD2 is the form which appears as a Clay Institute Millenium Prize problem [**Wi00**], but a number of the essential difficulties are already present in BSD1, particularly in the "if" statement.

1.5. Some known results

The analytic continuation of $L(E,s)$ is often stated separately from the Birch and Swinnerton-Dyer conjecture, as an important conjecture in its own right, which has now been proved (for elliptic curves over $\mathbb{Q}$) thanks to the work of Wiles and its subsequent refinements.

THEOREM 1.13. *The L-function $L(E,s)$ extends to an entire function on $\mathbb{C}$ and has a functional equation relating its value at s and $2-s$, of the form*

$$\Lambda(E,s) = \pm\Lambda(E,2-s),$$

where

$$\Lambda(E,s) := (2\pi)^{-s}\Gamma(s)N^{s/2}L(E,s),$$

and where

$$\Gamma(s) = \int_0^\infty e^{-t}t^{s-1}dt$$

is the Γ*-function.*

This theorem was proved by Wiles [**Wi95**] and Taylor-Wiles [**TW95**] in the case where E is a semistable elliptic curve; the full result was established after a series of strengthenings of Wiles' method [**Di96**], [**CDT99**], [**BCDT01**]. Some of the key concepts underlying the proof of Wiles' result – modular forms, and their associated L-series – are introduced in the next chapter.

While little is known in general about the Birch and Swinnerton-Dyer conjecture, one does have the following important partial result, which was originally proved *assuming* the validity of theorem 1.13, before that theorem was itself established. Reversing the chronological development, theorem 1.14 is stated here as an unconditional result.

THEOREM 1.14 (Gross-Zagier, Kolyvagin). *Let* E *be an elliptic curve over* $\mathbb{Q}$.

1. *If* $L(E,1) \neq 0$, *then* $\#E(\mathbb{Q}) < \infty$, *i.e.,* $r = 0$.
2. *If* $L(E,1) = 0$ *and* $L'(E,1) \neq 0$, *then* $r = 1$, *and there is an efficient method for calculating* $E(\mathbb{Q})$.

In both cases $\text{Ш}(E/\mathbb{Q})$ *is finite.*

Thus, both BSD2 of conjecture 1.12 and the Shafarevich-Tate conjecture are proved when $\mathrm{ord}_{s=1} L(E,s) \leq 1$.

One of the goals of this monograph is to discuss at least a few of the ideas that go into the proofs of theorem 1.13 and 1.14. Both rely crucially on the notion of *modularity* which is the focus of the next chapter.

FURTHER RESULTS AND REFERENCES

Tate's article [**Ta74**] presents what has by now become the standard account of the theory of elliptic curves. It forms the basis for both the classic graduate text by Silverman [**Si86**] and Husemoller's book [**Hu87**], and the less comprehensive but more elementary undergraduate treatise [**ST92**]. All three references contain a discussion of the proof of the Mordell-Weil theorem. Also recommended is the short book by Cassels [**Cas91**] which focusses on local-global principles and on the concepts surrounding the modern formulation of the proof of the Mordell-Weil theorem.

Beyond the questions arising from the lack of effectivity in the Mordell-Weil theorem, there are a number of other tantalising open problems suggested by this result. Foremost among these is the question of whether the rank of elliptic curves over the rationals can be arbitrarily large. Mestre [**Mes91**] has shown that there are infinitely many elliptic curves over $\mathbb{Q}$ with rank ≥ 12, and curves with rank as large as 21 have been found by computer search. There is no number field F for which the rank of $E(F)$ is known to get arbitrarily large as E varies over the elliptic curves defined over F, although such is expected to be the case for any number field.

There is an abundance of textbooks covering more advanced topics, such as [**Si94**] or Koblitz's book [**Kob93**] which explores the connections between the Birch and Swinnerton-Dyer conjecture and the ancient congruent number problem of determining which integers are the areas of right-angled triangles with rational sides lengths. Also of relevance to the topics of chapter 2 is the book of Knapp

[**Kn92**] emphasising the relation between the theory of elliptic curves and modular forms.

The conjecture of Birch and Swinnerton-Dyer was formulated in the early 60's in a series of two articles [**BSD63**] and [**BSD65**]. The precise relation (in which a surprising factor of $\sqrt{2}$ makes an appearance) between the asymptotics of the finite product occuring in the crude form of the Birch and Swinnerton-Dyer conjecture (conjecture 1.11) and the behaviour of $L(E, s)$ at $s = 1$ is discussed in [**Go**] and [**Con**].

Exhibiting many formal analogies with the case where F is a number field is the case where F is a function field - the field $k(T)$ of rational functions in an indeterminate T over a finite field k, or a finite extension of $k(T)$. While largely ignored in this monograph, the study of elliptic curves over function fields exhibits many similarities with that of elliptic curves over number fields, while at the same time appearing to be more tractable. For example, the Mordell-Weil theorem holds in this context, and it is also known that the rank of elliptic curves over $k(T)$ can become arbitrarily large, by work of Tate [**TS67**] and Ulmer [**Ul**]. (The latter work produces elliptic curves over $\mathbb{F}_p(T)$ of arbitrarily large rank whose j-invariants do not lie in the field of constants. For a variant of this construction based on the theory of Heegner points described in chapter 3, see [**Da03**].) More germane to this monograph, a lot more is known concerning the Birch and Swinnerton-Dyer conjecture for elliptic curves over function fields. The L-function $L(E/F, s)$ can be defined in this context, is known to have a functional equation, and its zeroes (of which there are finitely many) satisfy an analogue of the Riemann hypothesis. Furthermore, Tate has shown that the rank of $E(F)$ is at most the order of vanishing of $L(E, s)$ at $s = 1$, and that the full Birch and Swinnerton-Dyer conjecture, in its strongest form (BSD3) holds for E/F if $\text{Ш}(E/F)$ is finite. Together with theorem 1.14, the evidence arising from the analogy between function fields and number fields is probably the best evidence we dispose of at present for the validity of the Birch and Swinnerton-Dyer conjecture.

Exercises

(1) Prove the descent lemma (lemma 1.5) in the text.

(2) Let E be an elliptic curve over a field F. The goal of this exercise is to describe the connecting homomorphism $\delta : E(F)/nE(F) \longrightarrow H^1(F, E_n)$. Given a point $P \in E(F)$, choose $P' \in E(\bar{F})$ such that $nP' = P$, and set

$$z_{P'}(\sigma) = \sigma P' - P', \qquad \text{for } \sigma \in G_F.$$

 (a) Show that $z_{P'}$ belongs to the group $Z^1_{cont}(G_F, E_n)$ of continuous 1-cocycles on G_F with values in E_n.

 (b) Show that the image of zP' in $H^1(F, E_n)$ depends only on P, and not on the choice of P' that was made to define it. (Call c_P this image.)

 (c) Show that the assignment $P \mapsto c_P$ yields an injective group homomorphism

$$E(F)/nE(F) \longrightarrow H^1(F, E_n),$$

 and that $\delta(P) = c_P$.

(3) Assuming that $\text{Ш}(E/F)$ is finite, describe an effective procedure for computing $E(F)$. (I.e., an algorithm which given (E, F) as input is guaranteed to terminate after a finite amount of time with a list of generators for $E(F)$.)

(4) Let E be an elliptic curve over $\mathbb{R}$ and let $\Lambda_E \subset \mathbb{C}$ be a period lattice associated to a differential on E over $\mathbb{R}$. Show that $E(\mathbb{R})$ has two connected components if and only if the lattices

$$\Lambda_E \text{ and } (\Lambda_E \cap \mathbb{R}) + (\Lambda_E \cap i\mathbb{R})$$

are equal, and that otherwise the former contains the latter with index 2.

(5) Let E be an elliptic curve defined over a number field F of odd degree. If n is odd, show that $E_n(F)$ is cyclic.

(6) Let E be an elliptic curve over $\mathbb{Q}$ with minimal discriminant Δ, and let p be a prime of multiplicative reduction for E, so that $E_{/\mathbb{Q}_p}$ is endowed with the Tate uniformisation of equation (1.9). If ℓ is a prime which does not divide $\mathrm{ord}_p(\Delta)$, show that the image of $G_{\mathbb{Q}}$ in $\mathrm{Aut}(E_\ell) \simeq \mathbf{GL}_2(\mathbb{F}_\ell)$ contains a unipotent element (conjugate to a non-scalar upper triangular matrix with ones on the diagonal).

(7) Let E be an elliptic curve over a finite extension K of $\mathbb{Q}_p$. Let π be a uniformiser for K and let $k = \mathcal{O}_K/(\pi)$ denote the residue field. Assume that E has *good reduction* at π.

(a) Show that the kernel of the natural reduction map

$$E(K) \longrightarrow E(k)$$

is a pro-p group.

(b) Conclude that if p does not divide n, then the natural reduction map $E_n(\bar{K}) \longrightarrow E_n(\bar{k})$ is injective.

(c) Still assuming that $p \nmid n$, show that the action of the local Galois group G_K on E_n is *unramified*, i.e., that the inertia subgroup of G_K acts trivially on E_n.

(d) If E is an elliptic curve over a number field K, show that $\mathrm{Sel}_n(E/K)$ is contained in $H^1_{n\Delta}(K, E_n)$.

(8) Let K be a number field, d a fixed integer, and S a finite set of places of K. The Hermite-Minkowski theorem asserts that there are finitely many extensions of K with degree $\leq d$ and unramified outside S. Assuming this basic result, show that the group $H^1_{n\Delta}(K, E_n)$ occuring in the proof of proposition 1.7 is finite.

(9) Let $\mathcal{O}$ be a local ring with maximal ideal $\mathfrak{m}$ and residue characteristic different from 2 and 3. Show that the elliptic curve with equation

$$y^2 = x^3 + ax^2 + b, \text{ with } a \in \mathcal{O}^\times \text{ and } b \in \mathfrak{m},$$

has multiplicative reduction modulo $\mathfrak{m}$. Show that this reduction is split (resp. non-split) if and only if a is a square (resp. a non-square) in $\mathcal{O}^\times$.

(10) Using the Hasse bound $|a_p| < 2\sqrt{p}$, show that the Euler product defining $L(E, s)$ in equation (1.19) converges for $\mathrm{re}(s) > 3/2$ and that this convergence is uniform on compact subsets in this region.

CHAPTER 2

Modular forms

The only results of any depth that can be proved about the L-function $L(E, s)$ associated to an elliptic curve E (and hence, about the Birch and Swinnerton-Dyer conjecture formulated in the previous chapter) rely on knowing that E is *modular*. This chapter introduces this property and describes some of its important applications, most notably to the proof of theorem 1.13.

2.1. Modular forms

Let $\mathcal{H}$ be the Poincaré upper half-plane

$$\mathcal{H} = \{z \in \mathbb{C} \text{ such that } \mathrm{Im}(z) > 0\}. \tag{2.1}$$

The group $\mathbf{GL}_2^+(\mathbb{R})$ consisting of matrices with strictly positive determinant acts on $\mathcal{H}$ by Möbius transformations according to the rule

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tau = \frac{a\tau + b}{c\tau + d}. \tag{2.2}$$

In this way $\mathbf{GL}_2^+(\mathbb{R})$ acts on $\mathcal{H}$ by hyperbolic isometries preserving the line element $ds^2 = (dx^2 + dy^2)/y^2$, where $\tau = x + iy$. The group $\mathbf{SL}_2(\mathbb{Z})$ of matrices with integer coefficients and determinant 1 acts discretely on $\mathcal{H}$. (Cf. exercises 1 and 2.)

Let Γ be any finite index subgroup of $\mathbf{SL}_2(\mathbb{Z})$.

DEFINITION 2.1. A modular form of weight k on Γ is a holomorphic function $f : \mathcal{H} \longrightarrow \mathbb{C}$ such that

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

(2) For any $\gamma \in \mathbf{SL}_2(\mathbb{Z})$, the function

$$f_{|\gamma}(\tau) := (c\tau + d)^{-k} f(\gamma\tau)$$

can be written in the form

$$\sum_{n=0}^{\infty} a_n^{\gamma} q^{n/h}, \text{ where } q = e^{2\pi i\tau}, \tag{2.3}$$

and h is a positive integer, the *width* of the cusp $\gamma^{-1}\infty = \frac{-d}{c}$. The expression $\sum_{n=0}^{\infty} a_n^{\gamma} q^{n/h}$ depends only on $\frac{-d}{c}$ (up to multiplying $q^{1/h}$ by an h-th roots of unity) and is called the *Fourier expansion* of f at $\frac{-d}{c}$.

A modular form f on Γ is called a *cusp form* if $a_0^{\gamma} = 0$ for all γ. Denote by $S_k(\Gamma)$ the complex vector space of cusp forms of weight k for Γ. It is a fact (explained, for example, in [**Sh71**], theorem 2.24) that $S_k(\Gamma)$ is a finite-dimensional $\mathbb{C}$-vector space.

Of primary importance for elliptic curves – and the only case that will be treated in any detail in this chapter – is the case where $k = 2$ and where

$$\Gamma = \Gamma_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathbb{Z}) \text{ such that } N \text{ divides } c \right\}$$

is the so-called *Hecke congruence group* of level N for some positive integer N. Denote by $S_2(N) := S_2(\Gamma_0(N))$ the space of cusp forms of weight 2 on $\Gamma_0(N)$.

The quotient $\mathcal{H}/\Gamma_0(N)$ inherits from the complex structure on $\mathcal{H}$ the structure of a (non-compact) Riemann surface. It is useful to compactify $\mathcal{H}/\Gamma_0(N)$ by adjoining to it a finite set of *cusps* which corrrespond bijectively to the $\Gamma_0(N)$-orbits of points in $\mathbb{P}_1(\mathbb{Q})$, with an appropriate definition of the topology and complex structure in a neighbourhood of these cusps (discussed for example in [**Kn92**], p. 311 or [**Sh71**], ch. 1). In this way the quotient $\mathcal{H}^*/\Gamma_0(N)$ of the extended upper half-plane $\mathcal{H}^* := \mathcal{H} \cup \mathbb{P}_1(\mathbb{Q})$ by the action of $\Gamma_0(N)$ becomes a compact Riemann surface. Let $X_0(N)$ denote the projective algebraic curve (for the time being, over $\mathbb{C}$) whose complex points are identified with this Riemann surface. The assignment which to $f \in S_2(N)$ associates the expression

$$\omega_f := 2\pi i f(\tau) d\tau \tag{2.4}$$

identifies $S_2(N)$ with the space of holomorphic differential forms on $X_0(N)(\mathbb{C})$. Thus, by the Riemann-Roch theorem, the space $S_2(N)$ is a finite-dimensional complex vector space of dimension equal to the genus of $X_0(N)$. This fact allows the explicit determination of the dimension of $S_2(N)$ which is carried out in exercise 3.

2.2. Hecke operators

The vector space $S_2(N)$ is equipped with a non-degenerate Hermitian inner product

$$\langle f_1, f_2 \rangle = \int_{\mathcal{H}/\Gamma_0(N)} f_1(\tau) f_2(\bar{\tau}) dx dy, \tag{2.5}$$

known as the *Petersson scalar product.* It is also equipped with an action of certain *Hecke operators* T_p indexed by rational primes p and defined by the rules

$$T_p(f) = \begin{cases} \frac{1}{p}\sum_{i=0}^{p-1} f(\frac{\tau+i}{p}) + pf(p\tau), & \text{if } p \not| N \\ \frac{1}{p}\sum_{i=0}^{p-1} f(\frac{\tau+i}{p}), & \text{if } p|N. \end{cases} \tag{2.6}$$

These operators act linearly on $S_2(N)$, and their effect on the q expansions at ∞ (cf. exercise 4) is given by the following simple formulae:

$$T_p(f) = \begin{cases} \sum_{p|n} a_n q^{n/p} + p \sum a_n q^{pn} & \text{if } p \not| N; \\ \sum_{p|n} a_n q^{n/p} & \text{if } p|N. \end{cases} \tag{2.7}$$

It is convenient to extend the definition of the Hecke operators to operators T_n indexed by arbitrary positive integers n by equating the coefficient of n^{-s} in the identity of formal Dirichlet series

$$\sum_{n=1}^{\infty} T_n n^{-s} := \prod_{p\not| N} (1 - T_p p^{-s} + p^{1-2s})^{-1} \prod_{p|N} (1 - T_p p^{-s})^{-1}. \tag{2.8}$$

Let $\mathbb{T}$ be the commutative subalgebra of $\mathrm{End}_{\mathbb{C}}(S_2(N))$ generated over $\mathbb{Z}$ by the Hecke operators T_n, and let $\mathbb{T}^0$ denote the subalgebra generated only by those operators T_n with $(n, N) = 1$. Although presented as a ring with infinitely many generators, the algebras $\mathbb{T}$ and $\mathbb{T}^0$ are in fact finitely generated in a strong sense:

PROPOSITION 2.2. *The Hecke algebras $\mathbb{T}$ and $\mathbb{T}^0$ are finitely generated as $\mathbb{Z}$-modules.*

SKETCH OF PROOF. Let $V = \hom(S_2(N), \mathbb{C})$ denote the vector space dual of $S_2(N)$. By the theory of the Abel-Jacobi map (cf. for example [**Mu**], cor. 3.8) the integral homology $H^1(X_0(N)(\mathbb{C}), \mathbb{Z})$ embeds as a sublattice Λ of V by associating to a closed cycle c on $X_0(N)(\mathbb{C})$ the functional $\eta_c \in V$ defined by

$$\eta_c(f) = \int_c \omega_f.$$

The action of $\mathbb{T}$ on $S_2(N)$ induces an action of this algebra on V by duality, which leaves stable the lattice Λ (cf. [**Kn92**], props. 11.23 and 11.24). Hence $\mathbb{T}$ is a subalgebra of $\mathrm{End}_{\mathbb{Z}}(\Lambda)$. Since this latter ring is finitely generated as a $\mathbb{Z}$-module, the same must be true for $\mathbb{T}$ (and, a fortiori, $\mathbb{T}^0$). $\square$

It is worth trying to understand precisely what the rank of $\mathbb{T}$ as a $\mathbb{Z}$-module is. Let $g = \mathrm{genus}(X_0(N)) = \dim_{\mathbb{C}}(S_2(N))$.

PROPOSITION 2.3. *The rank of $\mathbb{T}$ is at most g.*

SKETCH OF PROOF. The action of complex conjugation τ on $X_0(N)(\mathbb{C})$ induces an action of τ on Λ which commutes with that of $\mathbb{T}$. Hence $\mathbb{T}$ preserves the submodules Λ^+ and Λ^- of Λ on which τ acts as multiplication by 1 and -1 respectively. Both Λ^+ and Λ^- are of rank g and hence $\mathbb{T}$ is identified with a commutative subalgebra of $M_g(\mathbb{Z})$ (after choosing a basis for Λ^+ or Λ^-). By exercise 5, there exists $T \in \mathbb{T}$ such that $\mathbb{T}$ contains $\mathbb{Z}[T]$ with finite index, and hence these two rings have the same rank as $\mathbb{Z}$-modules. On the other hand, $\mathbb{Z}[T]$ is generated by $1, T, \ldots, T^{g-1}$ since T satisfies its characteristic polynomial, and hence has rank at most g. $\square$

PROPOSITION 2.4. *The rank of $\mathbb{T}$ is equal to g.*

PROOF. Let $T_{\mathbb{C}} = \mathbb{T} \otimes \mathbb{C}$. There is a natural $\mathbb{C}$-bilinear pairing

$$\langle\ ,\ \rangle : T_{\mathbb{C}} \times S_2(N) \longrightarrow \mathbb{C}$$

given by $\langle T, f\rangle = a_1(Tf)$, where $a_1(g)$ denotes the first Fourier coefficient of the cusp form g. It follows from equation (2.7) (cf. exercise 4) that

$$\langle T_n, f\rangle = a_n(f). \tag{2.9}$$

Hence the pairing $\langle\ ,\ \rangle$ is non-degenerate on the right, so that the natural map

$$S_2(N) \longrightarrow \hom(T_{\mathbb{C}}, \mathbb{C})$$

induced by $\langle\ ,\ \rangle$ is injective. Hence $\dim_{\mathbb{C}}(T_{\mathbb{C}}) \geq g$, and therefore $\mathrm{rank}(\mathbb{T}) \geq g$. The result now follows from proposition 2.3. $\square$

The structure of $\mathbb{T}$ as a $\mathbb{Z}$-module leads to the following important fact about modular forms in $S_2(N)$ which shows the existence of many modular forms in this space having integer Fourier coefficients. The usefulness of modular forms in number theory stems from the fact that quite often these Fourier coefficients encode interesting arithmetic information.

COROLLARY 2.5. *The space $S_2(N)$ has a basis consisting of modular forms with integer Fourier coefficients.*

PROOF. By equation (2.9), the space of modular forms with integer Fourier coefficients is equal to the lattice dual to $\mathbb{T} \subset \mathbb{T}_{\mathbb{R}} = V^+$. It follows that this module has rank g over $\mathbb{Z}$. $\square$

Denote by $S_2(N, \mathbb{Z})$ and $S_2(N, \mathbb{R})$ the spaces of modular forms with integer and real Fourier coefficients respectively. Note that $S_2(N, \mathbb{Z})$ is a lattice in $S_2(N, \mathbb{R})$ and that

$$\mathrm{rank}_{\mathbb{Z}}(S_2(N, \mathbb{Z})) = \dim_{\mathbb{R}}(S_2(N, \mathbb{R})) = \dim_{\mathbb{C}}(S_2(N)) = g.$$

2.3. Atkin-Lehner theory

It is important to analyse the eigenspace decomposition of $S_2(N)$ under the action of the algebras of commuting operators $\mathbb{T}$ and $\mathbb{T}^0$. To begin, we have

LEMMA 2.6. *If T belongs to $\mathbb{T}^0$, then it is self-adjoint with respect to the Petersson scalar product.*

PROOF. See [**Kn92**] thm. 9.18 ad 8.22, or [**Og69**]. Better yet, work out exercise 6. $\square$

It follows from lemma 2.6 combined with the spectral theorem for commuting self-adjoint operators that $S_2(N)$ decomposes as an orthogonal direct sum of simultaneous eigenspaces for $\mathbb{T}^0$:

$$S_2(N) = \oplus_\lambda S_\lambda^0, \tag{2.10}$$

where the direct sum is taken over all $\mathbb{C}$-algebra homomorphisms $\lambda : \mathbb{T}^0 \longrightarrow \mathbb{C}$, and S_λ^0 denotes the corresponding eigenspace in $S_2(N)$. The eigenspaces S_λ^0 need not be one-dimensional (cf. exercise 7). On the other hand, if $\lambda : \mathbb{T} \longrightarrow \mathbb{C}$ is a ring homomorphism defined on the full Hecke algebra $\mathbb{T}$, and S_λ is its associated eigenspace, then one does have:

LEMMA 2.7 (Multiplicity one). *The eigenspace S_λ attached to $\lambda : \mathbb{T} \longrightarrow \mathbb{C}$ is one-dimensional.*

PROOF. This is because all the Fourier coefficients of a form $f \in S_\lambda$ are completely determined by $a_1(f)$, by the rule, immediate from (2.7)

$$a_n(f) = a_1(f)\lambda(T_n).$$

$\square$

As is shown in exercise 7, the space $S_2(N)$ does not decompose in general into a direct sum of the one-dimensional eigenspaces S_λ: the operators in $\mathbb{T}$ (unlike those in $\mathbb{T}^0$) need not act semisimply on $S_2(N)$. However, there is a distinguished subspace of $S_2(N)$, the so-called space of *newforms*, which decomposes as a direct sum of one-dimensional eigenspaces under both the actions of $\mathbb{T}$ and $\mathbb{T}^0$. More precisely, a modular form in $S_2(N)$ is said to be an *oldform* if it is a linear combination of functions of the form $f(d'z)$, with $f \in S^2(N/d)$ and $d'|d > 1$. The *new subspace* of $S_2(N)$, denoted $S_2^{\mathrm{new}}(N)$, is the orthogonal complement of the space $S_2^{\mathrm{old}}(N)$ of oldforms with respect to the Petersson scalar product.

THEOREM 2.8 (Atkin-Lehner). *The eigenspaces for the action of $\mathbb{T}^0$ on the space $S_2^{\mathrm{new}}(N)$ are all one-dimensional. The same assertion holds* a fortiori *for the action of $\mathbb{T}$ and we have an orthogonal decomposition:*

$$S_2(N) = S_2^{\mathrm{old}}(N) \oplus_\lambda \mathbb{C} f_\lambda,$$

where the sum is taken over all algebra homomorphisms $\lambda : \mathbb{T} \longrightarrow \mathbb{C}$ corresponding to eigenvectors in $S_2^{\mathrm{new}}(N)$, and $f_\lambda(\tau) = \sum_{n=1}^\infty \lambda(T_n) e^{2\pi i n \tau}$.

PROOF. See [**AL70**], or [**DI95**], sec. 6 for a survey. $\square$

The simultaneous eigenvector f_λ is sometimes called a *normalised eigenform* or simply a *newform* of level N. Note that it satisfies $a_1(f) = 1$.

2.4. L-series

To a newform f of level N is attached the L-series

$$L(f,s) = \sum_{n=1}^{\infty} a_n n^{-s},$$

where $a_n := a_n(f) = \lambda(T_n)$. This L-function enjoys the following important properties:

Euler product: It admits the Euler product factorisation:

$$L(f,s) = \prod_{p \nmid N} (1 - a_p p^{-s} + p^{1-2s})^{-1} \prod_{p | N} (1 - a_p p^{-s})^{-1},$$

as can be seen by applying λ to the formal identity (2.8) relating the Hecke operators T_n.

Integral representation: A direct calculation (cf. exercise 8) shows that

$$\Lambda(f,s) := (2\pi)^{-s}\Gamma(s) N^{s/2} L(f,s) = N^{s/2} \int_0^\infty f(it) t^{s-1} dt, \tag{2.11}$$

where $\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} dt$ is the Γ-function.

Functional equation: The involution w_N defined on $S_2^{\mathrm{new}}(N)$ by the rule

$$w_N(f) = \frac{-1}{N\tau^2} f\left(\frac{-1}{N\tau^2}\right) \tag{2.12}$$

commutes with the Hecke operators in $\mathbb{T}^0$ and hence preserves the eigenspaces S_λ. It follows that for any newform f of level N,

$$w_N(f) = \epsilon f, \quad \text{where } \epsilon = \pm 1. \tag{2.13}$$

The sign ϵ is an important invariant of f. Hecke showed (cf. exercise 9) that $L(f,s)$ satisfies the following functional equation:

$$\Lambda(f,s) = -\Lambda(w_N(f), 2-s) = -\epsilon \Lambda(f, 2-s). \tag{2.14}$$

2.5. Eichler-Shimura theory

The fact that the L-series $L(f,s)$ has properties identical to those that are known (such as the Euler product factorisation of equation (1.19)) or expected (such as the functional equation) for the L-series $L(E,s)$ attached to an elliptic curve in chapter 1 suggests that there might be a relationship between these two ostensibly different types of L-series. The following theorem of Eichler and Shimura establishes such a relationship in one direction.

THEOREM 2.9. *Let f be a normalised eigenform whose Fourier coefficients $a_n(f)$ are integers. Then there exists an elliptic curve E_f over $\mathbb{Q}$ such that*

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

BRIEF SKETCH OF PROOF. The crucial construction which to such a newform f associates the elliptic curve E_f was discovered by Shimura. A key fact underlying Shimura's construction is the fact that the algebraic curve $X_0(N)$ has a natural model defined over $\mathbb{Q}$. This curve, denoted $X_0(N)$ as before, is the solution to the (coarse) moduli problem of classifying pairs (A, C) where A is an elliptic curve and C is a cyclic subgroup of A of order N. The element $\tau \in \mathcal{H}/\Gamma_0(N)$ corresponds to the point in $X_0(N)(\mathbb{C})$ associated to the pair

$$(A_\tau, C_\tau) := \left(\mathbb{C}/\langle 1, \tau\rangle, \langle \frac{1}{N} \rangle\right), \tag{2.15}$$

and it can be checked that this correspondence is well-defined and sets up a bijection between points $\tau \in \mathcal{H}/\Gamma_0(N)$ and isomorphism classes of pairs (A, C) as above. An equation for $X_0(N)$ over $\mathbb{Q}$ can be written down using the modular function j defined using the j-invariant of equation (1.3) of chapter 1, by the rule:

$$j(\tau) = j(A_\tau).$$

The functions $j(\tau)$ and $j(N\tau)$ are related by a single polynomial relation $f_N(x,y)$ with coefficients in $\mathbb{Q}$ which yields a plane model for the algebraic curve $X_0(N)$, albeit one which is highly singular and not easily computable in practice except for small values of N. In this model, if F is any subfield of $\mathbb{C}$, and $\tau \in \mathcal{H}/\Gamma_0(N)$ corresponds to a point in $X_0(N)(F)$, then $(j(\tau), j(N\tau))$ belongs to F^2. The Hecke operators acting on $S_2(N)$ arise geometrically from certain correspondences (denoted $T_p \subset X_0(N) \times X_0(N)$ by abuse of notation). More precisely, the correspondence T_p is given as the locus of points in $X_0(N) \times X_0(N)$ attached to pairs which are related by a cyclic p-isogeny (defined as a cyclic p-isogeny of the underlying curves which induces an isomorphism between the level N structures.) Let $J_0(N)$ denote the Jacobian variety of $X_0(N)$; it is an abelian variety over $\mathbb{Q}$ of dimension $g = \text{genus}(X_0(N)) = \dim(S_2(N))$. The Hecke correspondences give rise to endomorphisms of $J_0(N)$ defined over $\mathbb{Q}$. Letting I_f denote the kernel of the ring homomorphism $\lambda : \mathbb{T} \longrightarrow \mathbb{Z}$ attached to f, the quotient $J_0(N)/I_f J_0(N)$ is the desired elliptic curve E_f.

The key issue which arises in showing the equality of L-functions $L(E_f, s)$ and $L(f,s)$ is to relate the coefficient $a_p(E)$ – obtained by counting points on E_f over $\mathbb{F}_p$, or, equivalently, as the trace of the Frobenius at p acting on the p-power division points of E_f – to the eigenvalue of the Hecke operators T_p. This relationship is established by using the Eichler–Shimura congruence relation satisfied by the Hecke correspondence T_p in characteristic p. For example, if p does not divide N

then the modular curve $X_0(N)$ has an integral model with good reduction at p (cf. the discussion in section 1.5 of [**DDT95**]) and one has

$$T_p = F + F^t \text{ on } X_0(N)_{/\mathbb{F}_p}$$

where F is the graph of the Frobenius morphism in characteristic p and F^t is its *transpose*. Deep results of Deligne and Carayol (cf. [**Car91**]) show that the level N of the newform f is equal to the arithmetically defined conductor of the elliptic curve E_f. A more complete discussion of the Eichler Shimura construction and the Eichler-Shimura congruence can be found in [**Sh71**], ch. 7, in [**Kn92**], chapter XI, or in section 1.7 of [**DDT95**]. □

The modular curve $X_0(N)$ is embedded in its Jacobian by sending a point P to the class of the degree 0 divisor $(P) - (i\infty)$. Let

$$\Phi_N : X_0(N) \longrightarrow E_f$$

be the *modular parametrisation* obtained by composing the embedding $X_0(N) \longrightarrow J_0(N)$ with the natural projection $J_0(N) \longrightarrow E_f$ arising from the Eichler-Shimura construction. The pullback $\Phi_N^*(\omega)$ of the Néron differential ω of E_f defined in chapter 1, (1.15), is a non-zero multiple of ω_f

$$\Phi_N^*(\omega) = c \cdot 2\pi i f(\tau) d\tau, \quad \text{with } c \in \mathbb{Q}^\times . \tag{2.16}$$

The rational number c is called the *Manin constant* attached to f. It is expected that $c = 1$ always, and this is known to be the case when N is square-free [**Ed89**].

For computational purposes, the following analytic description of the modular parametrisation

$$\Phi_N : \mathcal{H}/\Gamma_0(N) \longrightarrow E_f(\mathbb{C})$$

is useful.

PROPOSITION 2.10. *Let Λ_E be the Néron lattice of E and let c be the Manin constant attached to E_f. Let $\Phi_w : \mathbb{C}/\Lambda_E \longrightarrow E(\mathbb{C})$ be the Weierstrass uniformisation of equation (1.6) of chapter 1. Then one has*

$$\Phi_N(\tau) = \Phi_w(z_\tau), \quad \textit{where } z_\tau = c\int_{i\infty}^{\tau} 2\pi i f(z) dz = c\sum_{n=1}^{\infty} \frac{a_n}{n} q^n, \ \textit{with } q = e^{2\pi i \tau}.$$

PROOF. By definition of the Abel-Jacobi map and the projection $J_0(N) \longrightarrow E_f$, one sees that the image of the divisor $(\tau) - (i\infty)$ is

$$\Phi_w\left(\int_{\Phi_N(i\infty)}^{\Phi_N(\tau)} \omega\right) = \Phi_w\left(\int_{i\infty}^{\tau} \Phi_N^* \omega\right)$$

by the change of variables formula. The result follows from (2.16). □

The Eichler-Shimura construction yields elliptic curves E for which the analytic continuation and functional equation of $L(E,s)$ can be established. But it remains unclear, at this stage of the discussion, how special is the class of elliptic curves over $\mathbb{Q}$ that can be obtained from newforms with integer Fourier coefficients.

2.6. Wiles' theorem

In the 1960's, (through the work of a number of mathematicians, most prominently Shimura, Taniyama, and Weil) there gradually emerged the remarkable insight that in fact all elliptic curves (up to isogeny) should be obtainable from the Eichler-Shimura construction. The resulting conjecture, eventually known as the Shimura-Taniyama-Weil conjecture, is now a theorem thanks to a series of works building on a fundamental breakthrough of Wiles.

THEOREM 2.11. *Let E be an elliptic curve over $\mathbb{Q}$ of conductor N. Then there exists a newform $f \in S_2(N)$ such that*

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

Furthermore, E is isogenous to the elliptic curve E_f obtained from f via the Eichler-Shimura construction.

REMARK 2.12. An elliptic curve over $\mathbb{Q}$ is said to be *semistable* if its conductor is squarefree, so that it has either good or multiplicative reduction at all primes p. The articles [**Wi95**] and [**TW95**] give a proof of the Shimura-Taniyama-Weil conjecture for this class of elliptic curves. This work was later strengthened in a series of articles ([**Di96**], [**CDT99**]) culminating in a proof of the complete conjecture in [**BCDT01**].

REMARK 2.13. The distinguished elliptic curve E_f in the $\mathbb{Q}$-isogeny class of E is commonly called the *strong Weil curve* attached to E (or to f).

We will say little about the ideas that go into the proof of theorem 2.11, having already devoted a survey article [**DDT95**] to this topic. Let us simply point out the key role played in Wiles' approach by the continuous two-dimensional ℓ-adic representations of $G_{\mathbb{Q}} = \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ attached to elliptic curves, and to modular forms. This method produces a modular form f for which one has the equality of L-series $L(E,s) = L(E_f,s)$. The fact that E and E_f are isogenous over $\mathbb{Q}$ then follows from the isogeny conjecture proved by Faltings.

Of most relevance to the questions considered in this monograph are the following corollaries of Wiles' result:

COROLLARY 2.14. *The L-function $L(E,s)$ has an analytic continuation and an integral representation of the form*

$$(2\pi)^{-s}\Gamma(s)L(E,s) = \int_0^\infty f(it)t^{s-1}dt,$$

for some modular form $f \in S_2(\Gamma_0(N))$, and a functional equation as in theorem 1.13 of chapter 1.

Recall the sign $-\epsilon$ occuring in the functional equation of $L(E,s)$ given by (2.13) and (2.14). When E is associated to the modular form $f \in S_2(N)$, we will set

$$\mathrm{sign}(E,\mathbb{Q}) := -\epsilon, \tag{2.17}$$

Note that $L(E,s)$ vanishes to even (resp. odd) order at $s = 1$ when $\mathrm{sign}(E,\mathbb{Q}) = 1$ (resp. $\mathrm{sign}(E,\mathbb{Q}) = -1$).

The second corollary is the existence of an explicit complex uniformisation

$$\Phi_N : \mathcal{H}^*/\Gamma_0(N) \longrightarrow E(\mathbb{C}),$$

described in proposition 2.10. (Strictly speaking, one obtains it by composing the map of that proposition with a rational isogeny $E_f \longrightarrow E$.) The uniformisation

Φ_N will play a key role in the construction of certain algebraic points on E, known as *Heegner points*, which in turn are crucial in the proof of the theorem of Gross-Zagier-Kolyvagin. This key application of the modular parametrisation is discussed in chapter 3.

2.7. Modular symbols

Given a positive integer N, it is useful to be able to determine in practice an explicit basis of Hecke eigenforms in $S_2(N)$. (One natural application, arising from theorem 2.11, would be to list the elliptic curves over $\mathbb{Q}$ of a given conductor N, taken up to isogeny.)

A useful device in carrying out the computation of $S_2(N)$ is the notion of *modular symbols* first singled out and studied by Birch and Manin.

DEFINITION 2.15. Let A be an abelian group. An A-valued *modular symbol* is a function

$$m : \mathbb{P}_1(\mathbb{Q}) \times \mathbb{P}_1(\mathbb{Q}) \longrightarrow A, \text{ denoted } m(x,y) = m\{x \to y\},$$

satisfying

(1) $m\{x \to y\} = -m\{y \to x\}$, for all $x, y \in \mathbb{P}_1(\mathbb{Q})$.
(2) $m\{x \to y\} + m\{y \to z\} = m\{x \to z\}$, for all $x, y, z \in \mathbb{P}_1(\mathbb{Q})$.

Denote by $\mathcal{M}(A)$ the group of A-valued modular symbols, and simply by $\mathcal{M}$ the group of $\mathbb{C}$-valued modular symbols. The group $\mathbf{GL}_2(\mathbb{Q})$ acts naturally on $\mathcal{M}(A)$ by the rule

$$(\gamma m)\{x \to y\} := m\{\gamma^{-1}x \to \gamma^{-1}y\}.$$

Given $f \in S_2(N)$ with associated $\Gamma_0(N)$-invariant differential $\omega_f := 2\pi i f(\tau)d\tau$, let λ_f be the modular symbol defined by

$$\lambda_f\{x \to y\} := \int_x^y \omega_f. \tag{2.18}$$

Note that the cuspidality of f ensures that this integral converges. To compute it, one may choose $\tau \in \mathcal{H}$, and write

$$\int_x^y \omega_f = \int_x^\tau \omega_f + \int_\tau^y \omega_f.$$

If $x = \infty$, then the integral from x to τ can be evaluated by using the Fourier expansion of ω_f at ∞, by the rule

$$\int_\infty^\tau \omega_f = \sum_{n=1}^\infty \frac{a_n}{n} e^{2\pi i n\tau}.$$

Otherwise, choosing a matrix $\gamma \in \mathbf{SL}_2(\mathbb{Q})$ with $\gamma x = \infty$, one may reduce to the previous case by noting that

$$\int_x^\tau \omega_f = \int_\infty^{\gamma\tau} \omega_{f|\gamma},$$

and using the fourier expansion of $f|\gamma$ at ∞ to evaluate the latter integral.

The $\Gamma_0(N)$-invariance of ω_f implies that λ_f belongs to the subspace $\mathcal{M}^{\Gamma_0(N)}$ of $\Gamma_0(N)$-invariant modular symbols. This latter space is endowed with a natural action of the Hecke operators T_p defined, for $p \not| N$, by

$$T_p(m)\{x{\to}y\} = m\{px{\to}py\} + \sum_{j=0}^{p-1} m\left\{\frac{x+j}{p}{\to}\frac{y+j}{p}\right\}.$$

In this way the map λ which to f associates λ_f becomes a $\mathbb{C}$-linear Hecke-equivariant map from $S_2(N)$ to $\mathcal{M}^{\Gamma_0(N)}$.

LEMMA 2.16. *The map $\lambda : f \mapsto \lambda_f$ is injective.*

PROOF. Suppose that $\lambda_f = 0$, and let F be the holomorphic function on $\mathcal{H}^*$ defined by

$$F(\tau) = \int_\infty^\tau \omega_f.$$

This function is $\Gamma_0(N)$-invariant, since for all $\gamma \in \Gamma_0(N)$, one may choose any $x \in \mathbb{P}_1(\mathbb{Q})$ and note that

$$F(\gamma\tau) - F(\tau) = \int_\tau^{\gamma\tau} \omega_f = \int_x^{\gamma x} \omega_f = \lambda_f\{x{\to}\gamma x\} = 0.$$

Hence F corresponds to a holomorphic function on the compact Riemann surface $X_0(N)(\mathbb{C})$. It is therefore constant by Liouville's theorem. (In fact, it vanishes identically, since $F(i\infty) = 0$.) It follows that $dF = \omega_f = 0$. □

Recall that g is the genus of the modular curve $X_0(N)$, i.e., the complex dimension of $S_2(N)$. Let s denote the number of cusps on this modular curve.

LEMMA 2.17. *The space $\mathcal{M}^{\Gamma_0(N)}$ has dimension $2g + s - 1$.*

PROOF. Let $\mathcal{F}$ denote the space of $\mathbb{C}$-valued functions on $\mathbb{P}_1(\mathbb{Q})$, and let $d : \mathcal{F} \longrightarrow \mathcal{M}$ be the map defined by the rule

$$(df)\{x{\to}y\} = f(y) - f(x).$$

Clearly d is surjective and its kernel is the space of constant functions. Taking the $\Gamma = \Gamma_0(N)$-cohomology of the exact sequence

$$0 \longrightarrow \mathbb{C} \xrightarrow{i} \mathcal{F} \xrightarrow{d} \mathcal{M} \longrightarrow 0$$

yields the cohomology exact sequence

$$\mathbb{C} \xrightarrow{i_0} \mathcal{F}^\Gamma \xrightarrow{d} \mathcal{M}^\Gamma \longrightarrow H^1(\Gamma, \mathbb{C}) \xrightarrow{i_1} H^1(\Gamma, \mathcal{F}).$$

The cokernel of i_0 has dimension $s - 1$, while the kernel of i_1 is identified with $H^1(X_0(N), \mathbb{C}) \simeq \mathbb{C}^{2g}$. (Cf. exercise 10.) The result follows. □

It follows from lemma 2.17 that the map λ is not surjective. The failure of surjectivity can be traced to two sources:

1. Complex conjugation acts naturally on $\mathcal{M}^{\Gamma_0(N)}$, but the image of λ is not closed under this action. This is because

$$\bar{\lambda}_f = \lambda_{\bar{f}},$$

where $\lambda_{\bar{f}}$ denote the $\mathcal{M}$-symbol attached to the anti-holomorphic differential $\bar{\omega_f} = -2\pi i \bar{f}(\tau) d\bar{\tau}$. The homomorphism

$$\lambda : S_2(N) \oplus \overline{S_2(N)} \longrightarrow \mathcal{M}^{\Gamma_0(N)} \tag{2.19}$$

remains injective and its image has codimension $s - 1$.

2. A $\Gamma_0(N)$-invariant modular symbol m is said to be *Eisenstein* if there exists a $\Gamma_0(N)$-invariant function $M : \mathbb{P}_1(\mathbb{Q}) \longrightarrow \mathbb{C}$ such that

$$m\{x \rightarrow y\} = M(y) - M(x), \qquad \text{for all } x, y \in \mathbb{P}_1(\mathbb{Q}).$$

The space of Eisenstein modular symbols has dimension $s-1$. The space of Eisenstein symbols is linearly disjoint from the image of the map λ of equation (2.19).

LEMMA 2.18. *The space $\mathcal{M}^{\Gamma_0(N)}$ is spanned by the Eisenstein modular symbols together with the symbols of the form λ_f and $\bar{\lambda}_f$, where $f \in S_2(N)$.*

The details for this are worked out in exercise 11.

Thanks to lemma 2.18, to compute a basis of eigenforms for $S_2(N)$, it suffices to find a basis for the space $\mathcal{M}^{\Gamma_0(N)}$, discarding the Eisenstein modular symbols, and diagonalising the action of the Hecke operators on this space. This computation rests on the following simple observations:

First observation: We say that two elements $x = \frac{a}{b}$ and $y = \frac{c}{d}$ of $\mathbb{P}_1(\mathbb{Q})$ are *adjacent* if $ad - bc = \pm 1$. (It is assumed here that the fractions representing x and y are given in lowest terms, and the convention $\infty = \frac{1}{0}$ is adopted.) The first observation is that an $\mathcal{M}$-symbol is *completely determined* by its values on pairs (x, y) of adjacent elements of $\mathbb{P}_1(\mathbb{Q})$. This is because any two elements x, y can be realised as the extreme terms of a finite sequence $x, x_1, x_2, \ldots, x_n, y$, in which consecutive elements are adjacent. Clearly it is enough to show this when $x = \infty$, and for that, one may take $x_j = \frac{p_j}{q_j}$ to be the j-th convergent in the continued fraction expansion for y.

Second observation: The group $\Gamma_0(N)$ acts naturally on the set of ordered pairs of adjacent elements of $\mathbb{P}_1(\mathbb{Q})$, and there are finitely many orbits for this action. In fact, two pairs $(\frac{a}{b}, \frac{c}{d})$ and $(\frac{a'}{b'}, \frac{c'}{d'})$ with $ad - bc = a'd' - b'c' = 1$ belong to the same orbit precisely when the ratios b/d and b'/d' are equal in $\mathbb{P}_1(\mathbb{Z}/N\mathbb{Z})$. Therefore an element $m \in \mathcal{M}^{\Gamma_0(N)}$ is completely determined by the function

$$[\;]_m : \mathbb{P}_1(\mathbb{Z}/N\mathbb{Z}) \longrightarrow \mathbb{C}$$

defined by

$$[b : d]_m := m\left\{\frac{a}{b} \rightarrow \frac{c}{d}\right\}, \text{ with } ad - bc = 1.$$

This immediately gives a crude upper bound on the dimension of $\mathcal{M}^{\Gamma_0(N)}$: it is at most the cardinality of $\mathbb{P}_1(\mathbb{Z}/N\mathbb{Z})$.

Third observation: The functions of the form $[\;]_m$ on $\mathbb{P}_1(\mathbb{Z}/N\mathbb{Z})$, where m is a $\Gamma_0(N)$-invariant modular symbol, are not unrestricted but satisfy a collection of linear relations. For example, the axiom

$$m\left\{\frac{a}{b} \rightarrow \frac{c}{d}\right\} = -m\left\{\frac{-c}{-d} \rightarrow \frac{a}{b}\right\}$$

satisfied by modular symbols leads to the relation

$$[x]_m = -\left[\frac{-1}{x}\right]_m, \text{ for all } x \in \mathbb{P}_1(\mathbb{Z}/N\mathbb{Z}), \tag{2.20}$$

while the property

$$m\left\{\frac{a}{b} \rightarrow \frac{c}{d}\right\} = m\left\{\frac{a}{b} \rightarrow \frac{a+c}{b+d}\right\} + m\left\{\frac{a+c}{b+d} \rightarrow \frac{c}{d}\right\}$$

leads to the relation

$$[x]_m = \left[\frac{x}{x+1}\right]_m + [x+1]_m, \text{ for all } x \in \mathbb{P}_1(\mathbb{Z}/N\mathbb{Z}). \tag{2.21}$$

In fact, equations (2.20) and (2.21) represent a full set of linear relations satisfied by the functions of the form $x \mapsto [x]_m$. (Cf. exercise 12.)

These three observations can be combined into an explicit algorithm for finding all the Hecke eigenforms in $S_2(N)$ for a given level N, the details of whose implementation are discussed for example in [**Cr97**]. The reader wishing to acquire some familiarity with the modular symbol method may also find it helpful to work out exercises 15 and 16.

In other applications, such as those discussed in chapter 8, one may be given an elliptic curve E whose conductor N and associated L-series coefficients $a_n(E)$ can be computed (in any specified range) by examining the behaviour of E over the various completions of $\mathbb{Q}$. One may wish to parlay this knowledge into an explicit and efficient calculation of the modular symbol $\lambda_E = \lambda_f$ attached to f_E. The following theorem (a consequence of a more general result of Manin and Drinfeld) is useful for this task. Let Λ_E denote the so-called Néron lattice of E, generated by the periods of a Néron differential on E against the homology of E. Let t_E be the greatest common divisor of the integers $p + 1 - a_p(E)$, where p ranges over all primes which are congruent to 1 modulo N.

THEOREM 2.19. *The modular symbol λ_E takes values in a lattice Λ, which is contained in $\frac{1}{t_E}\Lambda_E$ with finite index.*

PROOF. If $x \in \mathbb{P}_1(\mathbb{Q})$ and $y = \gamma x$ are equivalent under the action of $\Gamma_0(N)$, then

$$\lambda_E\{x\to y\} = \int_x^y \omega_f = \int_x^{\gamma x} \omega_f = c^{-1} \int_{\Phi_N(x\to\gamma x)} \omega_E,$$

where $\Phi_N(x\to\gamma x)$ is the image under Φ_N of the closed path on $X_0(N)$ joining x to γx. Hence this expression belongs to Λ_E. In the general case, one exploits the action of the Hecke operators by noting that, for all $p \equiv 1 \pmod N$, the expression

$$(T_p - (p+1))\, \lambda_E\{x\to y\}$$

can be written as a sum of expressions of the form $\lambda_E\{x_j\to y_j\}$, where x_j and y_j belong to the same $\Gamma_0(N)$-orbit (Cf. exercise 13.) By varying the p, it follows that $t_E\lambda_E\{x\to y\}$ belongs to Λ_E, as was to be proved. $\square$

The space $\mathcal{M}^{\Gamma_0(N)}$ is equipped with a $\mathbb{C}$-linear involution W_∞ defined by

$$(W_\infty m)\{x\to y\} = m\{-x\to -y\}. \tag{2.22}$$

A modular symbol is called a plus (resp minus) symbol if it is in the corresponding eigenspace for W_∞. If f belongs to $S_2(N, \mathbb{R})$, then the action of W_∞ is related to that of complex conjugation on $\mathcal{M}$ by the rule:

$$(W_\infty \lambda_f)\{x\to y\} = \bar{\lambda}_f\{x\to y\}. \tag{2.23}$$

In particular, if f is an eigenform with integer Fourier coefficients, attached to an elliptic curve E, one may write

$$\lambda_E\{x\to y\} = \lambda_E^+\{x\to y\}\Omega^+ + \lambda_E^-\{x\to y\} i\Omega^-,$$

where Ω^+ (resp Ω^-) is the unique positive generator of the lattice in $\mathbb{R}$ gnerated by the real (resp imaginary) parts of elements of Λ. The functions $\lambda_E^\pm$ are integer-valued

$\Gamma_0(N)$-invariant modular symbols which are in the plus and minus eigenspace for W_∞ respectively.

Further results and references

Among the many textbooks that cover the theory of modular forms, [**Sh71**] and [**Og69**] are classics. Knapp's book [**Kn92**] on elliptic curves contains a fairly complete account of Eichler-Shimura Theory.

The Eichler-Shimura construction extends naturally to eigenforms f having non-rational Fourier coefficients. Since these coefficients are the eigenvalues of self-adjoint matrices with entries in $\mathbb{Z}$, they generate an order $\mathcal{O}$ in a totally real field K of degree $n \leq g$. The Eichler-Shimura construction, a direct generalisation of the construction in the proof of theorem 2.9, associates to f an abelian variety A_f of dimension n. The action of the Hecke operators on A_f yields a subring of $\mathrm{End}_{\mathbb{Q}}(A_f)$ isomorphic to $\mathcal{O}$, so that $\mathrm{End}_{\mathbb{Q}}(A_f) \otimes \mathbb{Q}$ contains a field K of degree $d = \dim(A_f)$. Abelian varieties with this property are said to be of $\mathbf{GL}_2$-type. It is conjectured that these are precisely (up to isogeny) the abelian varieties over $\mathbb{Q}$ that can occur as irreducible factors of $J_0(N)$, but one is still far from being able to prove this natural generalisation of the Shimura-Taniyama-Weil conjecture. See [**Ri94**] for a discussion of this and related conjectures.

The reader wishing to understand the proof of Wiles' theorem in the prototypical example of semi-stable elliptic curves can do no better than to consult the articles [**Wi95**] and [**TW95**] which contain the initial breakthroughs. A more pedagogically oriented exposition of these ideas is given in in [**DDT95**]. For a discussion of the subtle technical issues needed to extend Wiles' breakthrough to all elliptic curves over $\mathbb{Q}$, see [**Di96**], [**CDT99**] and [**BCDT01**].

Before it could be proved, the Shimura-Taniyama-Weil conjecture was supported by a certain amount of numerical evidence, and by Weil's converse to Hecke's theorem, asserting that if $L(E, s)$, as well as sufficiently many of its twists by Dirichlet characters, has a functional equation of the type described in theorem 1.13, then E is isogenous to an elliptic curve E_f obtained from an eigenform f via the Eichler-Shimura construction. A discussion of converse theorems can be found in [**Gel75**] for example.

The modular symbol method is very flexible and has been implemented by Cremona [**Cr97**] to draw up tables of elliptic curves of conductor ≤ 5000. A different method due to Mestre (cf. [**Mes91**] and [**Gr87**]) based on the so-called Brandt matrices associated to orders in definite quaternion algebras over $\mathbb{Q}$, is more restricted in scope (working well typically for elliptic curves of prime conductor) but appears to be more efficient than the modular symbol method in those situations where it can be readily applied.

Exercises

(1) Show that the action of a matrix $\gamma \in \mathbf{SL}_2(\mathbb{R})$ on $\mathcal{H}$ by Möbius transformations preserves the hyperbolic metric on $\mathcal{H}$.

(2) Let X and Y be Hausdorff topological spaces. A map $\pi : X \longrightarrow Y$ is called *proper* if $\pi^{-1}(K)$ is compact whenever K is compact, and *open* if it maps open subsets of X to open subsets of Y.

(a) Show that if $\pi : X \longrightarrow Y$ is open and proper, and $\Gamma \subset X$ is discrete, then $\pi(\Gamma)$ is discrete in Y.

(b) Let G be a locally compact Hausdorff topological group and let H be a compact subgroup. If Γ is a discrete subgroup of G, show that the natural image of Γ in G/H (in the natural topology on G/H induced by the topology on G) is discrete. Show that this ceases to be true in general if H is not assumed to be compact.

(c) Show that $\mathbf{PSL}_2(\mathbb{Z})$ acts discretely on $\mathcal{H}$ by Möbius transformations. (I.e., the orbits for this action are discrete in the usual Euclidean topology on $\mathcal{H}$.)

(3) Let $\Gamma = \mathbf{SL}_2(\mathbb{Z})$ and let

$$\pi : \mathcal{H}^*/\Gamma_0(N) \longrightarrow \mathcal{H}^*/\Gamma, \quad \Gamma_0(N)\tau \mapsto \Gamma\tau$$

be the natural (branched) covering map of Riemann surfaces. For $\tau \in \mathcal{H}^*$, let Γ_τ denote the stabiliser of τ in Γ and e_τ the index of $\Gamma_\tau \cap \Gamma_0(N)$ in Γ_τ. Let $\mathbb{P}_1(\mathbb{Z}/N\mathbb{Z})$ denote the projective line over the ring $\mathbb{Z}/N\mathbb{Z}$. It is equipped with a natural action of $\mathbf{SL}_2(\mathbb{Z})$.

(a) Show that e_τ is equal to the ramification index of π at $\Gamma_0(N)\tau$.

(b) Show that if τ is not equivalent under Γ to i, $\rho := e^{2\pi i/3}$, or ∞, then π is unramified at $\Gamma_0(N)\tau$, i.e., $e_\tau = 1$.

(c) Let n_i denote the number of fixed points of the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ acting on $\mathbb{P}_1(\mathbb{Z}/N\mathbb{Z})$. Show that π is unramified at n_i distinct points of $X_0(N)$ lying above $\tau = i$, and has ramification index 2 at the remaining points above i.

(d) Let n_ρ denote the number of fixed points of the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$ acting on $\mathbb{P}_1(\mathbb{Z}/N\mathbb{Z})$. Show that π is unramified at n_ρ distinct points of $X_0(N)$ lying above $\tau = \rho$, and has ramification index 3 at the remaining points above ρ.

(e) Show that the points in $\pi^{-1}(\infty)$ are in bijection with the orbits for the group $\Gamma_\infty = \langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \rangle$ acting on $\mathbb{P}_1(\mathbb{Z}/N\mathbb{Z})$, and that if the point P corresponds to the orbit $\mathcal{O}$, then $e_P = \#\mathcal{O}$.

(f) Using the Riemann-Hurwitz formula, give a formula for the genus of $X_0(N)$ (and hence, the dimension of $S_2(N)$) in terms of the orbit decompositions of $\mathbb{P}_1(\mathbb{Z}/N\mathbb{Z})$ under the actions of the three cyclic groups Γ_i, Γ_ρ, and Γ_∞. Make this formula explicit if N is a prime or a prime power.

(g) Show that there are no cusp forms of level N for $N \le 10$, and that $S_2(11)$ is one-dimensional.

(h) Use Wiles' theorem to conclude that there are no elliptic curves over $\mathbb{Q}$ of conductor < 11, and precisely one isogeny class of elliptic curves of conductor 11. How might one go about proving such a result, *without* using Wiles' theorem?

(4) Show that the function $T_p f$ described in formula (2.6) belongs to $S_2(N)$, so that T_p is indeed a well-defined linear endomorphism of $S_2(N)$. Verify formula (2.7) in the text decribing the effect of T_p on the Fourier expansion of f at ∞. Using this formula and the definition of the general Hecke operator T_n, prove equation (2.9) in the text.

(5) Let R be a commutative subring of $M_g(\mathbb{Z})$. Show that there exists $T \in R$ such that $\mathbb{Z}[T]$ is contained in R with finite index.
(6) Show that the Hecke operator T_p acting on $S_2(N)$ is self-adjoint with respect to the Petersson scalar product when $p \nmid N$.
(7) Let f be an eigenform of level N. Show that the modular forms $f(z)$, $f(pz)$ and $f(p^2z)$ belong to $S_2(Np^2)$ and are simultaneous eigenvectors for all the Hecke operators T_ℓ with ℓ not dividing Np^2. Show that the Hecke operator T_p preserves the space spanned by these three eigenforms, but that its action need not be diagonalisable.
(8) Check formula (2.11) in the text.
(9) Prove the functional equation (2.14) in the text.
(10) Complete the proof of lemma 2.17 by showing that the cokernel of i_0 and the kernel of i_1 are of dimension $s-1$ and $2g$ respectively.
(11) Supply the details in the proof of lemma 2.18.
(12) Show that equations (2.20) and (2.21) represent a full set of linear relations satisfied by the functions of the form $x \mapsto [x]_m$, where m is a $\Gamma_0(N)$-invariant modular symbol.
(13) Complete the proof of theorem 2.19 by showing that, for all $p \equiv 1 \pmod{N}$, the expression

$$(T_p - (p+1))\,\lambda_E\{x\to y\}$$

can be written as a sum of expressions of the form $\lambda_E\{x_j\to y_j\}$, where x_j and y_j belong to the same $\Gamma_0(N)$-orbit.
(14) Prove equation (2.23) relating the action of the involution W_∞ and complex conjugation acting on modular symbols of the form λ_f with $f \in S_2(N, \mathbb{R})$.
(15) Let $\Gamma = \Gamma_0(11)$ be the Hecke congruence group of level 11. Show that the space $\mathcal{M}^\Gamma$ of Γ-invariant modular symbols is a three-dimensional complex vector space, and write an explicit basis for these modular symbols consisting of eigenforms for the Hecke operators and for the involution W_∞. Use this to compute, for $p \le 7$, the number of points in $E(\mathbb{F}_p)$, for E any elliptic curve over $\mathbb{Q}$ of conductor 11 (which is unique up to isogeny, by exercise 3 (h).)
(16) Carry out exercise 15 with $N = 14$ or 37 instead of 11.

CHAPTER 3

Heegner points on $X_0(N)$

Chapter 2 explained how thanks to Wiles' theorem any elliptic curve E over $\mathbb{Q}$ of conductor N is attached to a normalised eigenform $f \in S_2(N)$, leading to the proof of the functional equation satisfied by $L(E, s)$ and the existence of the *modular parametrisation*

$$\Phi_N : \mathcal{H}^*/\Gamma_0(N) \longrightarrow E(\mathbb{C}).$$

The most important arithmetic application of this explicitly computable parametrisation arises through the theory of *complex multiplication.* This theory is of great importance in its own right, allowing the analytic construction, from values of modular functions evaluated at quadratic arguments, of class fields of imaginary quadratic fields. It can also be used, in conjunction with the parametrisation Φ_N, to exhibit algebraic points on E defined over such class fields. These points, known as *Heegner points*, are an essential ingredient in the proof of theorem 1.14 of Gross-Zagier and Kolyvagin.

3.1. Complex multiplication

Let $K \subset \mathbb{C}$ be a quadratic imaginary subfield of $\mathbb{C}$. We may write $K = \mathbb{Q}(\omega_D)$, where $D < 0$ is the discriminant of K and

$$\omega_D = \begin{cases} \frac{1+\sqrt{D}}{2} & \text{if } D \equiv 1 \pmod 4 \\ \frac{\sqrt{D}}{2} & \text{otherwise.} \end{cases}$$

It will be convenient to fix once and for all an embedding of an algebraic closure $\bar{K}$ of K into $\mathbb{C}$.

An *order* in K is a subring $\mathcal{O}$ of K which generates K as a $\mathbb{Q}$-vector space and is finitely generated as a $\mathbb{Z}$-module. Every order is contained in the *maximal order* $\mathcal{O}_K = \mathbb{Z}[\omega_D]$, and is uniquely determined by its *conductor* c, a positive non-zero integer such that

$$\mathcal{O} = \mathbb{Z} \oplus \mathbb{Z}c\omega_D.$$

Let $A = \mathbb{C}/\Lambda$ be an elliptic curve over $\mathbb{C}$. Its endomorphism ring is identified with

$$\mathrm{End}(A) = \{\alpha \in \mathbb{C} \text{ such that } \alpha\Lambda \subset \Lambda\}.$$

Hence it is isomorphic to a discrete subring of $\mathbb{C}$. Such a ring is isomorphic either to $\mathbb{Z}$, or to an order in a quadratic imaginary field K.

DEFINITION 3.1. An elliptic curve $A/\mathbb{C}$ is said to have *complex multiplication* if its endomorphism ring is isomorphic to an order in a quadratic imaginary field. More precisely, given such an order $\mathcal{O}$, one says that A has *complex multiplication* by $\mathcal{O}$ if $End(A) \simeq \mathcal{O}$.

If A has complex multiplication by $\mathcal{O}$, the corresponding period lattice of A is a projective $\mathcal{O}$-module of rank one, whose isomorphism class depends only on the isomorphism type of A. Conversely, if $\Lambda \subset \mathbb{C}$ is a projective $\mathcal{O}$-module of rank one, the corresponding elliptic curve $A = \mathbb{C}/\Lambda$ has complex multiplication by $\mathcal{O}$. Hence there is a bijection

$$\left\{ \begin{array}{l} \text{Elliptic curves with CM by } \mathcal{O}, \\ \text{up to isomorphism.} \end{array} \right\} \xrightarrow{\simeq} \left\{ \begin{array}{l} \text{Rank one projective } \mathcal{O}\text{-modules,} \\ \text{up to isomorphism.} \end{array} \right\}.$$

The set on the right is called the *Picard group*, or the *Class group*, of $\mathcal{O}$ and is denoted $\mathrm{Pic}(\mathcal{O})$. The group structure on it arises from the tensor product of modules. When $\mathcal{O} = \mathcal{O}_K$ is the maximal order, then $\mathrm{Pic}(\mathcal{O})$ is identified with the usual ideal class group of K. In any case the group $\mathrm{Pic}(\mathcal{O})$ is finite. (For more on the structure and size of the group $\mathrm{Pic}(\mathcal{O})$, see exercise 1.)

It follows from the finiteness of $\mathrm{Pic}(\mathcal{O})$ that there are finitely many isomorphism classes of elliptic curves with complex multiplication by $\mathcal{O}$. Let $\mathrm{Ell}(\mathcal{O})$ be the set of all such isomorphism classes, and let $A_1, \ldots, A_h$ be representatives for each class in $\mathrm{Ell}(\mathcal{O})$.

THEOREM 3.2. *The j-invariants $j(A_i)$ (with $1 \le i \le h$) are algebraic numbers.*

PROOF. Let $j = j(A_1)$. The elliptic curve $A = A_1$ is isomorphic over $\mathbb{C}$ to the curve with equation

$$y^2 + xy = x^3 + \frac{36}{j - 1728}x - \frac{1}{j - 1728} \tag{3.1}$$

Thus A can defined over $\mathbb{Q}(j)$, and in fact over the ring $\mathbb{Q}[j, 1/(j-1728)]$. If j is transcendental, this ring is contained in a field of transcendence degree one, and admits infinitely many distinct homomorphisms to $\mathbb{C}$. The corresponding specialisations of A would then yield infinitely many non-isomorphic elliptic curves with complex multiplication by $\mathcal{O}$, contradicting the finiteness of $\mathrm{Ell}(\mathcal{O})$. $\square$

We wish to gain a more precise understanding of the field generated by $j = j(A_1)$, i.e., the minimal field of definition of A_1, as well as the action of $\mathrm{Gal}(\bar{K}/K)$ on the algebraic numbers $j(A_j)$.

Note that previously $\mathcal{O} = \mathrm{End}(A)$ was identified with a subring of $\mathbb{C}$ by the rule

$$\alpha^*\omega_A = \alpha\Omega_A, \quad \text{for } \alpha \in \mathcal{O}, \tag{3.2}$$

where ω_A is any regular differential on A over $\mathbb{C}$. More generally, if A and its endomorphisms are defined over a field L, the purely algebraic condition (3.2) makes sense over L and singles out a ring homomorphism $\mathcal{O} \longrightarrow L$. In particular, any such L contains the fraction field K of $\mathcal{O}$, and an element of $\mathcal{O}$ gives rise in a natural way to an element of L via its effect on the cotangent space of A over L.

The collection $\mathrm{Ell}(\mathcal{O})$ is equipped with a natural simply transitive action of $\mathrm{Pic}(\mathcal{O})$ by the rule

$$[\Lambda] * [A] := \mathrm{Hom}(\Lambda, A), \tag{3.3}$$

(cf. [**Se67**], p. 294). More concretely, if $\mathfrak{p}$ is any prime ideal of K whose norm is prime to c, the inclusion $\mathfrak{p} \longrightarrow \mathcal{O}$ yields an isogeny

$$A = \mathrm{Hom}(\mathcal{O}, A) \longrightarrow \mathrm{Hom}(\mathfrak{p}, A),$$

whose kernel is identified with $\mathrm{Hom}(\mathcal{O}, A[\mathfrak{p}]) = A[\mathfrak{p}]$, where $A[\mathfrak{p}]$ denotes the subgroup scheme of elements in A which are annihilated by all elements of $\mathfrak{p}$. Thus

$$[\mathfrak{p}] * [A] = A/A[\mathfrak{p}]. \tag{3.4}$$

It follows directly from this description (cf. exercise 6) that the action of $\mathrm{Pic}(\mathcal{O})$ on $\mathrm{Ell}(\mathcal{O})$ commutes with the natural action of $G_K := \mathrm{Gal}(\bar{K}/K)$ on this set, so that the action of G_K on $\mathrm{Ell}(\mathcal{O})$ is encoded in a homomorphism

$$\eta : G_K \longrightarrow \mathrm{Pic}(\mathcal{O}), \quad \text{satisfying } A^\sigma = \eta(\sigma) * A, \text{ for all } \sigma \in G_K. \tag{3.5}$$

Note that by the commutativity of $\mathrm{Pic}(\mathcal{O})$, the definition of η does not depend on the choice of base curve A made to define it.

The mere existence of η is enough to show that the j-invariants $j(A_i)$ are defined over an abelian extension $H := \bar{K}^{\ker \eta}$ of K.

We now describe H more precisely in terms of class field theory.

Let

$$\mathbb{A}_{K,f} \subset \prod_{\ell \neq \infty} (K \otimes \mathbb{Q}_\ell)$$

denote the ring of finite adèles of K, and let

$$\hat{\mathcal{O}} := \prod_{\ell \neq \infty} (\mathcal{O} \otimes \mathbb{Z}_\ell)$$

be the closure of $\mathcal{O}$ in $\mathbb{A}_{K,f}$. For each prime λ of K, let K_λ denote the completion at the corresponding non-archimedean valuation. The group $K_\lambda^\times$ can be viewed naturally as a subgroup of the group $\mathbb{A}_{K,f}^\times$ of finite idèles attached to K. Let $\iota_\lambda(x)$ denote the idèle attached to $x \in K_\lambda^\times$. On the global level, K (resp. $K^\times$) can be viewed as a subring (resp. a subgroup) of $\mathbb{A}_{K,f}$ (resp $\mathbb{A}_{K,f}^\times$) via the natural diagonal embedding.

The group $\mathrm{Pic}(\mathcal{O})$ admits an adelic description, via the identification

$$\mathrm{Pic}(\mathcal{O}) = \mathbb{A}_{K,f}^\times / K^\times \hat{\mathcal{O}}^\times,$$

in which the class of the idèle α corresponds to the homothety class of the lattice $(\alpha^{-1}\hat{\mathcal{O}}) \cap K \subset \mathbb{C}$.

The following is a special case of the main theorem of class field theory (cf. for example [**CF67**], ch, VII, thm. 5.1):

THEOREM 3.3. *There exists an abelian extension H_c of K which is unramified outside of the primes dividing c, and whose Galois group is naturally identified, via the Artin map, with* $\mathrm{Pic}(\mathcal{O})$.

If $\mathfrak{p}$ is a prime ideal of K which is prime to c, we denote by $\pi_\mathfrak{p}$ a uniformiser of $K_\mathfrak{p}$, and by $[\mathfrak{p}]$ the class in $\mathrm{Pic}(\mathcal{O})$ attached to the finite idèle $\iota_\mathfrak{p}(\pi_\mathfrak{p})$. The Artin reciprocity law map

$$\mathrm{rec} : \mathrm{Pic}(\mathcal{O}) \longrightarrow \mathrm{Gal}(H_c/K)$$

sends the element $[\mathfrak{p}]$ to the inverse $\sigma_\mathfrak{p}^{-1}$ of the Frobenius element $\sigma_\mathfrak{p}$ at $\mathfrak{p}$.

The extension H_c whose existence is guaranteed by theorem 3.3 is called the *ring class field* of K attached to $\mathcal{O}$, or the *ring class field of K of conductor c*.

PROPOSITION 3.4. *The abelian extension H is equal to the ring class field H_c. More precisely, for all primes $\mathfrak{p}$ of K which do not divide c,*

$$\eta(\sigma_\mathfrak{p}) = [\mathfrak{p}] \in \mathrm{Pic}(\mathcal{O}).$$

PROOF. Fix an elliptic curve $A \in \mathrm{Ell}(\mathcal{O})$. Let Σ be the set of primes $\mathfrak{p}$ of $\mathcal{O}_K$ which satisfy the following conditions:

(1) $\mathfrak{p}$ is unramified in H/K;
(2) The curve A has good reduction at all the primes of H above $\mathfrak{p}$;
(3) The prime $\mathfrak{p}$ does not divide the norm (from H to K) of $j(A_s)-j(A_t)$ for all $s \neq t$;
(4) The norm of $\mathfrak{p}$ is a rational prime $p \in \mathbb{Z}$, i.e., p splits or ramifies in $K/\mathbb{Q}$.

Note that the set of primes of K satisfying these conditions has Dirichlet density one, and hence, by the Cebotarev density theorem, the corresponding Frobenius elements generate $\mathrm{Gal}(H/K)$. Let $\sigma_{\mathfrak{p}}$ be as before the Frobenius element in $\mathrm{Gal}(H/K)$ attached to the prime $\mathfrak{p} \in \Sigma$. Choose a prime $\mathfrak{p}'$ of H above $\mathfrak{p}$ and let $\bar{A}$ denote the reduction of A at $\mathfrak{p}'$. It is defined over a finite field $\mathbb{F}$ of characteristic p: the residue field of H at $\mathfrak{p}'$. We make two key observations (whose proofs are worked out in detail in exercise 7):

(1) The elliptic curve $\bar{A}$ is *ordinary*, i.e., there is (up to composition with an automorphism of A) a unique inseparable isogeny of degree p from $\bar{A}$, given by the Frobenius morphism:
$$\mathrm{Frob} : \bar{A} \longrightarrow \bar{A}^p,$$
where $\bar{A}^p$ denotes the elliptic curve over $\mathbb{F}$ obtained from A by applying the Frobenius map (raising to the pth power) to the coefficients of a defining equation for $\bar{A}$, and Frob is the algebraic morphism of degree p sending (x, y) to (x^p, y^p).
(2) The elliptic curve $A/A[\mathfrak{p}]$ is defined over H, and the natural projection
$$A \longrightarrow A/A[\mathfrak{p}]$$
is a purely inseparable morphism modulo $\mathfrak{p}'$.

It follows from these two observations that $A/A[\mathfrak{p}] = [\mathfrak{p}] * A$ is congruent, modulo $\mathfrak{p}'$, to $\sigma_{\mathfrak{p}}(A) \equiv \bar{A}^p$. Since all the A_j are distinct modulo $\mathfrak{p}'$, it follows that $\eta(\sigma_{\mathfrak{p}}) = [\mathfrak{p}]$. Since this formula holds on a set of primes of K of density one, it then holds for all the primes of K which are unramified in H; in particular $H = H_c$. $\square$

Let $M_2(\mathbb{Z})$ be the algebra of 2×2 matrices with entries in $\mathbb{Z}$. Given any $\tau \in \mathcal{H}$, the *order* associated to τ is defined to be
$$\mathcal{O}_\tau := \{\gamma \in M_2(\mathbb{Z}) \text{ such that } \det\gamma \neq 0 \text{ and } \gamma\tau = \tau\} \cup \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}.$$
The elements of $\mathcal{O}_\tau$ consist precisely of the matrices in $M_2(\mathbb{Z})$ which have both the column vectors $\begin{pmatrix} \tau \\ 1 \end{pmatrix}$ and $\begin{pmatrix} \bar{\tau} \\ 1 \end{pmatrix}$ as eigenvectors. It is transparent from this description that $\mathcal{O}_\tau$ is closed under both addition and multiplication and that it is a commutative subring of $M_2(\mathbb{Z})$. The natural map which to $\gamma \in \mathcal{O}_\tau$ associates the complex number z_γ satisying
$$\gamma \begin{pmatrix} \tau \\ 1 \end{pmatrix} = z_\gamma \begin{pmatrix} \tau \\ 1 \end{pmatrix}. \tag{3.6}$$
also allows $\mathcal{O}_\tau$ to be viewed as a discrete subring of $\mathbb{C}$. Note further that he order $\mathcal{O}_\tau$ is isomorphic to the endomorphism ring of the elliptic curve $A_\tau = \mathbb{C}/\langle 1, \tau\rangle$. This is because, for any $\gamma \in \mathcal{O}_\tau$, multiplication by the complex number z_γ of equation

(3.6) preserves the lattice $\langle\tau, 1\rangle$ and hence induces a complex endomorphism m_γ of A_τ. In fact the map $\gamma \mapsto m_\gamma$ identifies $\mathcal{O}_\tau$ with $\mathrm{End}_{\mathbb{C}}(A_\tau)$.

If $\mathcal{O}$ is any order in a quadratic imaginary field $K \subset \mathbb{C}$, we can write

$$CM(\mathcal{O}) = \{\tau \in \mathcal{H}/\mathbf{SL}_2(\mathbb{Z}) \text{ such that } \mathcal{O}_\tau = \mathcal{O}\}.$$

The class group $\mathrm{Pic}(\mathcal{O})$ acts on $CM(\mathcal{O})$ as follows: any class $\alpha \in \mathrm{Pic}(\mathcal{O})$ can be represented by an integral ideal $I \subset \mathcal{O}$ such that $\mathcal{O}/I$ is cyclic. Choose such an I. The lattice $(1, \tau)I^{-1}$ is a projective $\mathcal{O}$-module which contains 1 as an indivisible element. Hence we can write

$$\langle 1, \tau\rangle I^{-1} = \langle 1, \tau'\rangle, \tag{3.7}$$

in which the generator τ' is well-defined modulo the action of $\mathbf{SL}_2(\mathbb{Z})$, and define

$$\alpha \star \tau := \tau'. \tag{3.8}$$

It is a routine matter (cf. exercise 5) to check that this rule does indeed endow $CM(\mathcal{O})$ with an action of $\mathrm{Pic}(\mathcal{O})$ which is compatible with the action of this group on $\mathrm{Ell}(\mathcal{O})$ introduced earlier.

We may thus reformulate the main theorem of complex multiplication in a purely analytic way, as follows:

THEOREM 3.5. *Let $K \subset \mathbb{C}$ be a quadratic imaginary field and let $\tau \in \mathcal{H} \cap K$ be an element of $\mathcal{H}$ which is quadratic over $\mathbb{Q}$. Then $j(\tau)$ belongs to H, where H is the ring class field attached to the order $\mathcal{O} = \mathcal{O}_\tau$. More precisely, for all $\alpha \in \mathrm{Pic}(\mathcal{O})$, and $\tau \in CM(\mathcal{O})$,*

$$j(\alpha \star \tau) = \mathrm{rec}(\alpha)^{-1} j(\tau).$$

3.2. Heegner points

Let N be a fixed positive integer and let $M_0(N)$ be the ring of 2×2 matrices with entries in $\mathbb{Z}$ which are upper-triangular modulo N, so that $\Gamma_0(N)$ is the group of units of determinant 1 in this ring. Adapting the terminology of the previous section, we now define the *associated order* of τ (relative to the level N) to be

$$\mathcal{O}_\tau^{(N)} := \{\gamma \in M_0(N) \text{ such that } \gamma\tau = \tau\} \cup \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}.$$

It is easy to see that

$$\mathcal{O}_\tau^{(N)} = \mathcal{O}_\tau \cap \mathcal{O}_{N\tau}.$$

The reader will note that the map from $\mathcal{H}$ to $E(\mathbb{C})$ induced by the modular parametrisation Φ_N is transcendental, since it is of infinite degree. Hence Φ_N is not normally expected to take on algebraic values when evaluated on algebraic arguments. The following theorem provides an important exception to this rule.

THEOREM 3.6. *Let τ be any element in $\mathcal{H} \cap K$, let $\mathcal{O} = \mathcal{O}_\tau^{(N)}$ be its associated order in $M_0(N)$, and let H/K be the ring class field attached to $\mathcal{O}$. Then $\Phi_N(\tau)$ belongs to $E(H)$.*

PROOF. By theorem 3.5, both $j(\tau)$ and $j(N\tau)$ belong to a ring class field H of K associated to the order $\mathcal{O}_\tau \cap \mathcal{O}_{N\tau}$. Hence $\Phi_N(\tau)$ is the image of a point in $X_0(N)(H)$ (its coordinates are given by $(J(\tau), j(N\tau))$ in the singular plane model of $X_0(N)$ given by the Nth modular polynomial) by the modular parametrisation Φ_N. Hence $\Phi_N(\tau)$ belongs to $E(H)$, since the map $X_0(N) \longrightarrow E$ induced by Φ_N is a map of algebraic curves defined over $\mathbb{Q}$. (Cf. the remarks in the proof of theorem 2.9 of chapter 2.) $\square$

Abusing notation, write $\mathcal{O}_\tau$ instead of $\mathcal{O}_\tau^{(N)}$ (trusting that the context will make it clear what level N is being used.) If $\mathcal{O}$ is any order in a quadratic imaginary field $K \subset \mathbb{C}$, we similarly rewrite

$$CM(\mathcal{O}) = \{\tau \in \mathcal{H}/\Gamma_0(N) \text{ such that } \mathcal{O}_\tau = \mathcal{O}\}.$$

We then lift the action of $\mathrm{Pic}(\mathcal{O})$ to $CM(\mathcal{O}) \subset \mathcal{H}/\Gamma_0(N)$ by setting $\alpha \star_N \tau := \tau'$, where $\tau' \in \mathcal{H}$ is chosen so that

$$\alpha \star \tau = \tau', \quad \alpha \star (N\tau) = N\tau' \qquad (\mathrm{mod}\ \mathbf{SL}_2(\mathbb{Z})).$$

Note that this property determines τ' modulo the action of $\Gamma_0(N)$. It is a routine matter (cf. exercise 5) to check that this rule gives a concrete description of the action of $\mathrm{Pic}(\mathcal{O})$ on $CM(\mathcal{O})$ which can thus be related to the the action of $\mathrm{Gal}(H/K)$ on $E(H)$ via the reciprocity law of class field theory.

THEOREM 3.7 (Shimura reciprocity law). *If τ belongs to $CM(\mathcal{O})$ and α belongs to* $\mathrm{Pic}(\mathcal{O})$*, then*

$$\Phi_N(\alpha \star \tau) = \mathrm{rec}(\alpha^{-1})\Phi_N(\tau).$$

PROOF. This follows directly from the reciprocity law of theorem 3.5. $\square$

3.3. Numerical examples

One of the charms of theorems 3.6 and 3.7 is that they lend themselves to concrete calculations and allow the construction, by analytic means, of a large supply of algebraic points on E defined over ring class fields of quadratic imaginary fields.

For example, one knows from the tables of [**Cr97**] (cf. also exercise 3 of chapter 2) that the elliptic curve of smallest conductor $N = 11$ is given by the equation

$$y^2 + y = x^3 - x^2 - 10x - 20, \tag{3.9}$$

while the imaginary quadratic order of smallest discriminant which embeds in $M_0(11)$ is $\mathcal{O}_K = \mathbb{Z}(\frac{1+\sqrt{-7}}{2})$. The field K has class number one, and the order

$$\mathcal{O} = \mathbb{Z} + \mathbb{Z}\begin{pmatrix} -4 & -2 \\ 11 & 5 \end{pmatrix}$$

is an order in $M_0(11)$ which is isomorphic to $\mathcal{O}_K$. (This order is unique, up to conjugation by the normaliser $\tilde{\Gamma}_0(11)$ of $\Gamma_0(11)$ in $\mathbf{PGL}(\mathbb{Q})$.) The fixed point τ for this order is

$$\tau = \frac{-9 + \sqrt{-7}}{22}.$$

The Fourier coefficients $a_n(E)$ of the modular form f can be calculated by counting points of E mod p, or by using the identity

$$f = \sum_{n=1}^{\infty} a_n q^n = q \prod_{n=1}^{\infty} (1 - q^{11n})^2 (1 - q^n)^2 = q - 2q^2 - q_2^3 q^4 + q^5 + \cdots$$

(The calculation of these coefficients a_n, given an equation for E, is a built-in feature of many symbolic algebra packages such as Pari, and calculating the first

1000 coefficients takes less than a second on a small computer.) Setting $q = e^{2\pi i\tau}$ and computing the image of

$$z = \sum_{n=1}^{1000} \frac{a_n}{n} q^n$$

in $E(\mathbb{C})$ under the Weierstrass uniformisation yields a point which agrees with the point

$$(x,y) = (\frac{1-\sqrt{-7}}{2}, -2-2\sqrt{-7})$$

to 35 decimal digits of accuracy. Further calculations of this kind are suggested in the exercises.

3.4. Heegner systems

PROPOSITION 3.8. *Let $\mathcal{O}$ be an order of discriminant prime to N. Then the set $CM(\mathcal{O})$ is non-empty if and only if all the primes dividing N split in $K/\mathbb{Q}$.*

PROOF. If $CM(\mathcal{O})$ is non-empty, then $\mathcal{O}$ can be realised as a subring of $M_0(N)$. Therefore, there is a ring homomorphism $\mathcal{O} \longrightarrow \mathbb{Z}/N\mathbb{Z}$. Since the conductor of $\mathcal{O}$ is assumed to be prime to N, it follows that all ℓ dividing N are split in K. $\square$

Because of this proposition, it is natural to require that the following *Heegner hypothesis* be satisfied

HYPOTHESIS 3.9. *All primes ℓ dividing N are split in $K/\mathbb{Q}$.*

Let n be any integer prime to N and let $\mathcal{O}_n$ be the order of K of conductor n. A point of the form $\Phi_N(\tau)$, with $\tau \in CM(\mathcal{O}_n)$, is called a *Heegner point of conductor n*. Let $HP(n) \subset E(H_n)$ denote the set of all Heegner points of conductor n in $E(H_n)$, where H_n denotes the ring class field of K of conductor n. The points in $HP(n)$ are related by the following norm-compatibilities:

PROPOSITION 3.10. *Let n be an integer and ℓ a prime number which are prime to N. Let $P_{n\ell}$ be any point in $HP(n\ell)$. Then there exist points $P_n \in HP(n)$ and (when $\ell|n$) $P_{n/\ell} \in HP(n/\ell)$ such that*

$$\mathrm{Trace}_{H_{n\ell}/H_n}(P_{n\ell}) = \begin{cases} a_\ell P_n & \textit{if } \ell \not| n \textit{ is inert in } K; \\ (a_\ell - \sigma_\lambda - \sigma_\lambda^{-1})P_n & \textit{if } \ell = \lambda\bar{\lambda} \not| n \textit{ is split in } K; \\ (a_\ell - \sigma_\lambda)P_n & \textit{if } \ell = \lambda^2 \textit{ is ramified in } K \\ a_\ell P_n - P_{n/\ell} & \textit{if } \ell | n. \end{cases}$$

PROOF. We content ourselves with the proof of the second norm-compatiblity relation, as the others are similar, but somewhat simpler. Let $(A \longrightarrow A')$ be the pair of N-isogenous elliptic curves corresponding to the point P_n. If $\ell = \lambda\bar{\lambda}$ is a prime of $\mathbb{Q}$ which is split in K and prime to N, then the action of $\mathrm{Gal}(\bar{K}/H_n)$ on $A[\ell]$ leaves invariant two cyclic subgroups of order ℓ: the groups $C_0 = A[\lambda]$, and $C_\infty = A[\bar{\lambda}]$, and permutes the remaining $\ell - 1$ subgroups $C_1, \ldots, C_{\ell-1}$ transitively. In fact, this permutation action factors through a simply transitive action of $\mathrm{Gal}(H_{n\ell}/H_n)$ on $\{C_1, \ldots, C_{\ell-1}\}$. Let $P_{n\ell}^{(j)}$ be the point in $E(H_{n\ell})$ corresponding to the pair $(A/C_j \xrightarrow{\varphi} A'/\varphi(C_j))$, and set $P_{n\ell} = P_{n\ell}^{(1)}$. The description of the Hecke operator T_ℓ in terms of cyclic ℓ-isogenies makes it apparent that

$$a_\ell P_n = P_{n\ell}^{(0)} + P_{n\ell}^{(\infty)} + P_{n\ell}^{(1)} + \cdots + P_{n\ell}^{(\ell-1)}. \tag{3.10}$$

On the other hand,

$$(3.11)\quad P_{n\ell}^{(0)} = \sigma_\lambda P_n, \quad P_{n\ell}^{(\infty)} = \sigma_\lambda^{-1} P_n, \quad P_{n\ell}^{(1)} + \cdots + P_{n\ell}^{(\ell-1)} = \mathrm{Trace}_{H_{n\ell}/H_n}(P_{n\ell}).$$

The result follows immediately from (3.10) and (3.11). For more details and a discussion of the remaining cases see [**Gr89**]. □

An element $\tau \in \mathrm{Gal}(H/\mathbb{Q})$ is called a *reflection* if its restriction to K is not the identity. Because of the dihedral nature of $\mathrm{Gal}(H/\mathbb{Q})$, any reflection is of order 2, and any two reflections differ by multiplication by an element of $\mathrm{Gal}(H/K)$. We will need the following behaviour of the points P_n under the action of a reflection.

PROPOSITION 3.11. *Let $\tau \in \mathrm{Gal}(H/\mathbb{Q})$ be a reflection. Then there exists $\sigma \in \mathrm{Gal}(H/K)$ such that*

$$\tau P_n = -\mathrm{sign}(E, \mathbb{Q}) \sigma P_n \pmod{E(H)_{\mathrm{tors}}},$$

where $\mathrm{sign}(E, \mathbb{Q})$ *is the sign attached to $E/\mathbb{Q}$ described in equation (2.17) of chapter 2.*

PROOF. See [**Gr84**]. □

DEFINITION 3.12. A *Heegner system* attached to (E, K) is a collection of points $P_n \in E(H_n)$ indexed by integers n prime to N, and satisfying the norm compatibility properties of proposition 3.10 together with the behaviour under the action of reflections described in proposition 3.11.

A Heegner system is said to be non-trivial if at least one of the points P_n is non-torsion.

THEOREM 3.13. *If (E, K) satisfies the Heegner hypothesis, then there is a non-trivial Heegner system attached to (E, K).*

PROOF. The union of the points $CM(n)$, as n ranges over all integers prime to N, is dense in $\mathcal{H}$ with respect to the complex topology, as soon as the Heegner hypothesis, which ensures that the sets $CM(n)$ are non-empty, is satisfied. Hence the image of these points in $E(\mathbb{C})$ is dense with respect to the complex topology, and, in particular, infinite. Let H_∞ denote the union of all the ring class fields of conductor prime to N. To rule out the possibility that all the points P_n are torsion, one uses the following lemma:

LEMMA 3.14. *The torsion subgroup of $E(H_\infty)$ is finite.*

PROOF. Any rational prime which is inert in K splits completely or is ramified in all ring class fields, so that the residue field of H_∞ at such a prime q is the field $\mathbb{F}_{q^2}$ with q^2 elements. Since the prime-to-q torsion in $E(H_\infty)$ injects into the finite group $E(\mathbb{F}_{q^2})$, it follows that the full torsion subgroup of $E(H_\infty)$ injects into $E(\mathbb{F}_{q_1^2}) \oplus E(\mathbb{F}_{q_2^2})$, where q_1 and q_2 are two distinct rational primes which are inert in K. □

It follows from lemma 3.14 and the discussion preceeding it that at least one (in fact, infinitely many) of the points $\Phi_N(\tau)$ is of infinite order in $E(H_\infty)$, proving theorem 3.13. □

The existence of a systematic supply of points defined over the ring class fields of an imaginary quadratic field represents in itself a noteworthy occurence. The next section examines the compatibility between this phenomenon and the conjecture of Birch and Swinnerton-Dyer.

3.5. Relation with the Birch and Swinnerton-Dyer conjecture

Let K be any number field, and let D_K be its discriminant. If v is a fractional ideal of K, denote by $|v| \in \mathbb{Q}$ the norm of v. If E is an elliptic curve over $\mathbb{Q}$ as before, one can consider as in the remarks closing chapter 1 the L-function $L(E/K, s)$ of E over K. This L-series can be expressed as an Euler product taken over the primes of K:

$$L(E/K, s) = \prod_v L_v(E/K, s),$$

where $L_v(E/K, s)^{-1}$ is a polynomial in $|v|^{-s}$ of degree at most 2 given by the rule

$$L_v(E/K, s) = \begin{cases} (1 - a_{|v|}|v|^{-s} + |v|^{1-2s})^{-1} & \text{if } v \not| N; \\ (1 - a_{|v|}|v|^{-s})^{-1} & \text{if } v|N. \end{cases}$$

Assume from now on that K is a *quadratic field*, so that one has (cf exercise 15)

$$L(E/K, s) = L(E, s)L(E', s) \tag{3.12}$$

where E' is the quadratic twist of E over K.

More generally, let

$$\chi : \mathrm{Gal}(H/K) \longrightarrow \mathbb{C}^\times$$

be any character of the ring class field H of conductor c with $(c, N) = 1$, and set $D = D_K c^2$. Define the twisted L-series by the rule

$$L(E/K, \chi, s) = \prod_v L_v(E/K, \chi, s),$$

where $L_v(E/K, \chi, s)$ is given, for $v \not| ND$, by the formula

$$L_v(E/K, \chi, s) = (1 - \chi(\sigma_v)a_{|v|}|v|^{-s} + \chi(\sigma_v)^2|v|^{1-2s})^{-1}. \tag{3.13}$$

(For the general formula describing the Euler factor for $v|ND$, see [**Gr84**], ch. III, sec. 19, or exercise 16.) It is useful to complete the definition of the local L-factor $L_v(E/K, \chi, s)$ to the infinite primes, by setting

$$L_\infty(E/K, \chi, s) = (2\pi)^{-2s}\Gamma(s)^2.$$

Let $A = N^2D^2/\gcd(N, D)$.

THEOREM 3.15. *Let*

$$\Lambda(E/K, \chi, s) = A^{s/2}L_\infty(E/K, \chi, s)L(E/K, \chi, s).$$

The L-function $L(E/K, \chi, s)$ has an analytic continuation to the entire complex plane, and satisfies a functional equation of the form

$$\Lambda(E/K, \chi, s) = \mathrm{sign}(E, K)\Lambda(E/K, \chi, 2 - s), \tag{3.14}$$

where $\mathrm{sign}(E, K) = \pm 1$ *is a sign which depends only on E and K, not on the ring class character χ of conductor prime to N.*

REMARK ON THE PROOF. The proof of the analytic continuation and functional equation for $L(E/K,s)$ and $L(E/K,\chi,s)$ relies on Rankin's method and its extensions by Jacquet [**Ja72**]. When K is an imaginary quadratic field, it is explained in detail in chapter IV of [**GZ84**]. See also the remarks in [**Gr84**], ch. III, sec. 21. □

While the proof of theorem 3.15 does not matter very much for our discussion, the shape of the functional equation in this theorem – in particular, the independence on χ of the sign in equation (3.14) – has a number of striking arithmetic consequences when combined with the Birch and Swinnerton-Dyer conjecture.

For example, if H is any ring class field of K of conductor prime to N, then by exercise 17, the L-function of E/H factors as

$$L(E/H,s)=\prod_{\chi} L(E/K,\chi,s), \tag{3.15}$$

where the product is taken over all the characters of $\mathrm{Gal}(H/K)$. It follows from theorem 3.15 that, if $\mathrm{sign}(E,K)=-1$,

$$L(E/K,\chi,s)=0 \quad \text{for all characters } \chi:\mathrm{Gal}(H/K)\longrightarrow\mathbb{C}. \tag{3.16}$$

Hence, by (3.15)

$$\mathrm{ord}_{s=1}L(E/H,s)\geq[H:K]. \tag{3.17}$$

The Birch and Swinnerton-Dyer conjecture then leads to the expectation that

$$\mathrm{rank}(E(H))\overset{?}{\geq}[H:K]. \tag{3.18}$$

The following conjecture is motivated by this predicted inequality:

CONJECTURE 3.16. *If* $\mathrm{sign}(E,K)=-1$*, then there is a non-trivial Heegner system attached to* (E,K)*.*

We turn to a description of $\mathrm{sign}(E,K)$ in the special case where E is *semistable* over K. The *analytic set* attached to E over K is the set of places of K which are archimedean or at which the curve E has split multiplicative reduction. (The reason for this terminology is that $S_{E,K}$ consists precisely of the places of K for which $E(K_v)$ admits an analytic uniformisation, given by the Weierstrass and Tate theory mentioned in chapter 1 when v is archimedean and non-archimedean respectively.) The following makes theorem 3.15 more precise:

THEOREM 3.17. *Suppose that* E *is an elliptic curve over* $\mathbb{Q}$ *and* K *is a quadratic field. Then*

$$\mathrm{sign}(E,K)=(-1)^{\#S_{E,K}}.$$

IDEA OF PROOF. The functional equation for $L(E/K,\chi,s)$ of [**Ja72**] expresses the sign in the functional equation as a product of local signs $\mathrm{sign}_v(E,K)$ indexed by the places v of K. It turns out that $\mathrm{sign}_v(E,K)=1$ if v does not belong to $S_{E,K}$, and is equal to -1 otherwise. A complete and explicit description of the "local root numbers" attached to the functional equations of L-series of elliptic curves is given in [**Ro96**]. □

REMARK 3.18. Much of the discussion above can be transposed to the setting where $\mathbb{Q}$ is replaced by an arbitrary number field F (and where E is an elliptic curve over F, while K is a quadratic extension of F). In that setting the L-function $L(E/K,\chi,s)$ is still expected to satisfy a functional equation of a shape similar to that given by theorem 3.15, with $\mathrm{sign}(E,K)$ depending only on E and K, not on the ring class character χ of conductor prime to the conductor of E. On the basis of the Birch and Swinnerton-Dyer conjecture, it is natural to make the following conjecture:

CONJECTURE 3.19. *Suppose that E is a semistable elliptic curve over a number field F and K is a quadratic extension of F. Let*

$$\mathrm{sign}(E,K) := (-1)^{\#S_{E,K}}.$$

If $\mathrm{sign}(E,K) = -1$, then there is a non-trivial Heegner system attached to (E,K).

Even though the statement of conjecture 3.19 is elementary and does not involve the notion of modularity, one knows at present of no method for tackling it directly without exploiting a connection between elliptic curves and automorphic forms. In fact, only in the rather limited number of cases where one can establish the analytic continuation and functional equation of $L(E/K,\chi,s)$ – by relating it to the L-series of an automorphic form on $\mathbf{GL}_2(F)$, as in the case $F = \mathbb{Q}$ covered by Wiles' theory – does one have any means at present of relating $\mathrm{sign}(E,K)$ to the behaviour of this associated L-series.

To illustrate how theorem 3.17 can be applied, suppose that E is a semistable elliptic curve over $\mathbb{Q}$ and that K is a quadratic imaginary field satisfying the Heegner hypothesis with respect to E. In that case

$$S_{E,K} = \{\lambda \text{ such that } \lambda|\ell|N \text{ and } E/\mathbb{Q}_\ell \text{ has split multiplicative reduction}\} \cup \{\infty\}.$$

Since each rational prime ℓ dividing N splits in K, the primes of K for which E has split multiplicative reduction come in pairs and hence $\#S_{E,K}$ is odd, so that $\mathrm{sign}(E,K) = -1$. Hence the Heegner point construction supplies a proof of conjecture 3.16 in the special case where (E,K) satisfies the Heegner hypothesis (cf. thm.3.13).

3.6. The Gross-Zagier formula

As in the previous sections, let E be an elliptic curve over $\mathbb{Q}$ and let K be an imaginary quadratic field such that (E,K) satisfies the Heegner hypothesis. Denote by $\{P_n\} = \{\Phi_N(\tau_n)\}$ the Heegner system arising from the points in $HP(n)$. Let

$$P_K = \mathrm{Trace}_{H_1/K}(P_1) \in E(K)$$

be the trace of a Heegner point of conductor 1 defined over the Hilbert class field of K. More generally, if $\chi : \mathrm{Gal}(H_n/K) \longrightarrow \mathbb{C}^\times$ is any primitive character of a ring class field extension of K of conductor n, let

$$P_n^\chi = \sum_{\sigma\in\mathrm{Gal}(H_n/K)} \bar{\chi}(\sigma) P_n^\sigma \in E(H_n) \otimes \mathbb{C}.$$

The following result provides the essential bridge between the Heegner system $\{P_n\}$ and the special values of the complex L-series $L(E/K,s)$ and its twists.

THEOREM 3.20 (Gross-Zagier, Zhang). *Let $\langle\ ,\ \rangle_n$ denote the canonical Néron-Tate height on $E(H_n)$ extended by linearity to a Hermitian pairing on $E(H_n)\otimes\mathbb{C}$. Then*

(1) $\langle P_K, P_K\rangle \doteq L'(E/K,1)$;
(2) $\langle P_n^\chi, P_n^{\bar\chi}\rangle \doteq L'(E/K,\chi,1)$.

Here the symbol $\doteq$ denotes equality up to a non-zero fudge factor, which can in principle be made explicit. Part 1 of theorem 3.20 was proved in [**GZ84**] (cf. thm. 2.1 of section V.2) and part 2 is proved in [**Zh01b**]. The main consequence of theorem 3.20 is that the Heegner vector P_n^χ is non-zero if and only if $L'(E/K,\chi,1)$ does not vanish.

3.7. Kolyvagin's theorem

A non-trivial Heegner system obviously yields certain lower bounds on the size of the Mordell-Weil group of E over ring class fields of K. The following theorem reveals that a Heegner system also leads (somewhat surprisingly, at first sight) to *upper* bounds on the Mordell-Weil group and the Shafarevich-Tate group of E/K.

THEOREM 3.21 (Kolyvagin). *Let $\{P_n\}$ be a Heegner system attached to (E,K). If P_K is non-torsion, then*

(1) *The Mordell-Weil group $E(K)$ is of rank one, so that P_K generates a finite-index subgroup of $E(K)$;*
(2) *The Shafarevich-Tate group of E/K is finite.*

The proof of this theorem is explained in chapter 10. (The interested reader may immediately skip to this chapter which is independent of the material in chapters 4–9.)

3.8. Proof of the Gross-Zagier-Kolyvagin theorem

We now explain how the results of sections 3.6 and 3.7 can be combined to prove theorem 1.14 of chapter 1. Recall that this theorem states:

THEOREM 3.22. *If E is an elliptic curve over $\mathbb{Q}$ and $\mathrm{ord}_{s=1}L(E,s)\leq 1$, then*

$$\mathrm{rank}(E(\mathbb{Q})) = \mathrm{ord}_{s=1}L(E,s) \quad \textit{and } \#Ш(E/\mathbb{Q}) < \infty.$$

SKETCH OF PROOF. Recall that $\mathrm{sign}(E,\mathbb{Q})$ denotes the sign in the functional equation for $L(E,s)=L(E/\mathbb{Q},s)$. Suppose first that $\mathrm{sign}(E,\mathbb{Q})$ is equal to -1. By a result of [**Wa85**] (see also [**MM97**]), there exist infinitely many quadratic Dirichlet characters ε such that

(1) $\varepsilon(\ell)=1$ for all $\ell|N$;
(2) $\varepsilon(-1)=-1$;
(3) $L(E,\varepsilon,1)\neq 0$.

Note that for all characters ε satisfying conditions 1 and 2 above, $L(E,\varepsilon,s)$ vanishes to even order at $s=1$. This is because the quadratic imaginary field corresponding to ε satisfies the Heegner hypothesis with respect to E, so that $L(E/K,s)=L(E,s)L(E,\varepsilon,s)$ vanishes to odd order at $s=1$. If $\mathrm{sign}(E,\mathbb{Q})$ is equal to 1, then for parity reasons $L(E,\varepsilon,1)=0$ for all quadratic Dirichlet characters satisfying conditions 1 and 2 above. In this case one invokes analytic results of [**BFH90**] and [**MM91**] which guarantee the existence of such a character ε for which

$$L'(E,\varepsilon,1)\neq 0.$$

In either case, let K be the quadratic imaginary field associated to ε. By construction,

(1) K satisfies the Heegner hypothesis relative to E;
(2) $\mathrm{ord}_{s=1}L(E/K,s)=1$, so that $L'(E/K,1)\neq 0$.

Let $\{P_n\}$ be the Heegner system arising from the CM points on $X_0(N)$ attached to K. The Gross-Zagier theorem implies that this Heegner system in non-trivial in the strong sense that P_K is non-torsion. Kolyvagin's theorem then implies that $E(K)$ has rank one, so that the quotient of $E(K)$ by $\langle P_K \rangle$ is finite, as is the Shafarevich-Tate group of E/K. By proposition 3.11, P_K belongs to $E(\mathbb{Q})$ (up to torsion) if and only if $\mathrm{sign}(E, \mathbb{Q}) = -1$. It follows that the rank of $E(\mathbb{Q})$ is equal to the order of vanishing of $L(E, s)$ at $s = 1$, as predicted by the Birch and Swinnerton-Dyer conjecture. Finally, the finiteness of $Ш(E/K)$ directly implies the finiteness of $Ш(E/\mathbb{Q})$ since the natural map $Ш(E/\mathbb{Q}) \longrightarrow Ш(E/K)$ induced by restriction has finite kernel. (Cf. exercise 18.) $\square$

One cannot emphasize enough the crucial role played in this proof by the Heegner system arising from CM points on $X_0(N)$.

Heegner systems are interesting objects in their own right, even beyond their striking application to the arithmetic of elliptic curves arising from the theorems of Gross-Zagier and Kolyvagin. It is therefore natural to examine the following question:

QUESTION 3.23. *Let E be an elliptic curve over $\mathbb{Q}$ and let K be a quadratic field with* $\mathrm{sign}(E, K) = -1$. *Is it possible to construct a non-trivial Heegner system attached to (E, K)?*

The next chapter presents the necessary background to give an essentially complete affirmative answer to this question (and thereby prove conjecture 3.16) in the case where K is an *imaginary* quadratic field. The case where K is real quadratic is more mysterious: one knows of no method for analytically constructing the class fields of real quadratic fields. In this case one must content oneself with a conjectural construction that works in certain special cases. The tools and concepts needed for this construction are introduced gradually in chapters 5, 6 and 7 and the construction is described in chapter 8.

Further results

An excellent introduction to the theory of complex multiplication is given in Serre's article on complex multiplication (chapter XII of [**CF67**]). A more lengthy and leisurely introduction, containing a wealth of historical and extra material, is the book [**Cox89**] by Cox.

The seminal article [**GZ84**] and the follow-up article [**GKZ87**] provide a rich source of information on Heegner points, and their connections with special values of the associated Rankin L-series. Other useful background on this topic can be found in [**Gr84**], [**Za85**], [**Zh01a**] and [**Zh01b**].

A useful exposition of the proof of Kolyvagin's theorem (which is also the focus of chapter 10) is given in [**Gr89**].

Exercises

(1) Let K be an imaginary quadratic field, and let $\mathcal{O} \subset \mathcal{O}_K$ be the order in K of conductor $c > 1$. If $h = \#\mathrm{Pic}(\mathcal{O}_K)$ is the class number of K, show that

$$\#\mathrm{Pic}(\mathcal{O}) = h \cdot \# \frac{(\mathcal{O}_K/c\mathcal{O}_K)^\times}{(\mathbb{Z}/c\mathbb{Z})^\times \mathcal{O}_K^\times}.$$

For $K = \mathbb{Q}(\omega_D)$ of class number one, attach to each element $\alpha \in (\mathcal{O}_K/c\mathcal{O}_K)^\times$ a projective $\mathcal{O}$-module Λ_α, in such a way that Λ_α and Λ_β are isomorphic if and only if $\alpha = \beta u$ for some $u \in (\mathbb{Z}/c\mathbb{Z})^\times \mathcal{O}_K^\times$. Compute $\mathrm{Pic}(\mathcal{O})$ for
(a) $\mathcal{O} = \mathbb{Z}[i]$, $\mathbb{Z}[2i]$, $\mathbb{Z}[3i]$, and $\mathbb{Z}[4i]$;
(b) $\mathcal{O} = \mathbb{Z}[\rho]$, $\mathbb{Z}[2\rho]$, and $\mathbb{Z}[3\rho]$, where $\rho = \omega_{-3}$.
(c) $\mathcal{O} = \mathbb{Z}[\omega_{-7}]$ and $\mathbb{Z}[11\omega_{-7}]$.

(2) Let E be an elliptic curve with complex multiplication by K. Assume that both E and the endomorphisms of E are defined over a number field H.
(a) Show that the natural image of $\mathrm{Gal}(\bar{H}/H)$ in $\mathrm{Aut}(E_n)$, for any integer n, is abelian.
(b) Conclude that E does not acquire multiplicative reduction over any finite extension of H. (Hint: use exercise 6 of chapter 1.)
(c) It is know that the prime ideals dividing the denominator of the j-invariant of E are precisely the ones at which E has potentially multiplicative reduction. Conclude that if τ belongs to $\mathcal{H} \cap K$, then $j(\tau)$ is an algebraic integer.

(3) Show that $j(\frac{1+i\sqrt{163}}{2})$ is an integer. Compute this integer using a symbolic algebra package such as PARI. Explain the curious identity

$$e^{\pi\sqrt{163}} = 262537412640768743.9999999996...$$

(4) Complete the proof of theorem 3.6 in the case where the conductor of $\mathcal{O}$ is not assumed to be prime to N.

(5) Check that formula (3.8) does yield a well-defined action of $\mathrm{Pic}(\mathcal{O})$ on $CM(\mathcal{O})$.

(6) Check that the action of $\mathrm{Pic}(\mathcal{O})$ on $\mathrm{Ell}(\mathcal{O})$ commutes with the natural action of $G_K := \mathrm{Gal}(\bar{K}/K)$ on this set.

(7) Let E be an elliptic curve with complex multiplication by an order $\mathcal{O}$ in a quadratic imaginary field K, defined over an abelian extension H of K. Let $\mathfrak{p}$ be a prime of K of norm a rational prime p, which is unramified in H/K and modulo which E and all its Galois conjugates have good reduction. Choose a prime $\mathfrak{p}'$ of H above $\mathfrak{p}$. Let $\bar{E}$ be the curve obtained from E by reducing modulo $\mathfrak{p}'$.
(a) Show that there is a unique (up to composition with automorphisms of the image) inseparable isogeny from $\bar{E}$ of degree p. (Hint: show that the isogeny $\bar{E} \longrightarrow \bar{E}/\bar{E}[\bar{\mathfrak{p}}]$ is separable, by choosing an ideal I of norm prime to p such that $\bar{\mathfrak{p}}I = (\lambda)$ is principal. Then show that the composed map $\bar{E} \longrightarrow \bar{E}/\bar{E}[\bar{\mathfrak{p}}] \longrightarrow \bar{E}/\bar{E}[\lambda] = \bar{E}$ is separable, by examining its effect on the tangent space of E, using equation (3.2).)
(b) By a similar argument, show that the natural projection $\bar{E} \longrightarrow \bar{E}/\bar{E}[\mathfrak{p}]$ is purely inseparable of degree p.

(8) For the following elliptic curves E and fundamental discriminants D, compute the class number h of D and a system $(x_1, y_1), \ldots, (x_h, y_h)$ of complex points approximating the Heeger points on E attached to $\mathcal{O}$ with an accuracy of at least 20 decimal digits. Decribe the polynomial of degree h with coordinates in $\mathbb{Q}$ satisfied by the x_j and the y_j.
(a) $E : y^2 + y = x^3 - x^2 - 10x - 20$, of conductor 11; with $D = -8, -19, -24$, or -43.
(b) $E : y^2 + xy + y = x^3 + 4x - 6$, of conductor 14; with $D = -31$ or -68.
(c) $E : y^2 + y = x^3 - x$, of conductor 37; with $D = -3, -4, -11$, or -67,

(9) Let E be the elliptic curve $y^2 = 4x^3 - 28x + 25$ of conductor 5077.
   (a) Show (using PARI or another symbolic algebra package, say) that the sign in the functional equation of $L(E,s)$ is -1 so that this L-series vanishes to odd order.
   (b) By a direct numerical calculation, verify that the Heegner point associated to the order of discriminant -4 is torsion.
   (c) Using the Gross-Zagier formula, conclude that $L(E,s)$ has a zero of order at least 3 at $s = 1$.

   The existence of such L-series with zeroes of high order, implied by the conjecture of Birch and Swinnerton-Dyer, plays a key role in Goldfeld's effective solution of the Gauss class number problem. It was not known unconditionally before the work of Gross and Zagier.

(10) Show that a CM point on $\mathcal{H}/\Gamma_0(N)$ corresponds to an integral point on the open modular curve $Y_0(N)$. (Hint: use the fact that an elliptic curve with complex multiplication has potentially good reduction at all primes.) Conclude that if the modular parametrisation Φ_N maps only the cusps of $X_0(N)$ to the origin of E, then the curve E has a Heegner system $\{P_n\}$ consisting entirely of integral points (relative to the minimal Weierstrass model for E).

(11) Prove a converse to the theorem of the previous exercise: the Heegner system of points $\{P_n\} = \{\Phi_n(\tau_n)\}$ consists entirely of integral points on E, if and only if the inverse image of the origin of E under Φ_N consists entirely of cusps on $X_0(N)$.

(12) Show that the integrality property of Heegner points is satisfied by the elliptic curves of conductor < 37 and the curve labelled $37A$ in Cremona's tables, but not by the curve labelled $37B$ in these tables.

(13) Produce an example of a Heegner point on the curve $37B$ which is not integral. (I am grateful to Antoine Gournay for carrying out the computer calculations necessary to formulate this exercise and the previous one.)

(14) * Give a complete list of the elliptic curves satisying the integrality property of Heegner points. (A useful reference to get started on this as yet unsolved problem is [**MSw-D74**].)

(15) Prove formula (3.12) in the text.

(16) The local Euler factor at p in the L-function of $E/\mathbb{Q}$ can be written as

$$(1 - \alpha_p p^{-s})(1 - \alpha'_p p^{-s}),$$

   where, for $p \not| N$, we have $\alpha_p + \alpha'_p = a_p(E)$, and $\alpha_p \alpha'_p = p$. Given a ring class character $\chi : \mathrm{Gal}(H/K) \longrightarrow \mathbb{C}^\times$, set

$$(\beta_p, \beta'_p) = \begin{cases} (\chi(\sigma_{\mathfrak{p}}), \chi(\sigma_{\mathfrak{p}})^{-1}) & \text{if } p = \mathfrak{p}\bar{\mathfrak{p}} \text{ splits in } K; \\ (1, -1) & \text{if } p \text{ is inert in } K; \\ (\chi(\sigma_{\mathfrak{p}}), 0) & \text{if } p = \mathfrak{p}^2 \text{ is ramified in } K \text{ but not in H}; \\ (0,0) & \text{if } p \text{ is ramified in } H/K. \end{cases}$$

   (a) Let V be the two-dimensional Artin representation of $G_{\mathbb{Q}}$, given by inducing χ from G_k to $G_{\mathbb{Q}}$. Show that the Artin L-series $L(V,s)$ is equal to

$$L(V,s) = \prod_p (1 - \beta_p p^{-s})^{-1}(1 - \beta'_p p^{-s})^{-1}.$$

(b) Show that when p does not divide N or the discriminant of H, the degree four Euler factor

$$(1-\alpha_p\beta_p p^{-s})(1-\alpha_p\beta_p' p^{-s})(1-\alpha_p'\beta_p p^{-s})(1-\alpha_p'\beta_p' p^{-s})$$

is equal to the product of the Euler factors in equation 3.13, taken over the primes v which divide p.

(c) Define the local factor at the rational prime p for $L(E/K,\chi,s)$ to be

$$(1-\alpha_p\beta_p p^{-s})(1-\alpha_p\beta_p' p^{-s})(1-\alpha_p'\beta_p p^{-s})(1-\alpha_p'\beta_p' p^{-s}).$$

Compute this Euler factor at the primes which divide N or the discriminant of H.

(17) Prove equation (3.15) in the text.

(18) Show that the natural map $\text{Ш}(E/\mathbb{Q}) \longrightarrow \text{Ш}(E/K)$ induced by restriction has finite kernel.

CHAPTER 4

Heegner points on Shimura curves

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

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

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 a quadratic imaginary 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 $P_n = \Phi_N(\tau_n)$ 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 repertoire 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. The standard example is the ring $M_2(F)$ of 2×2 matrices with entries in F. 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 adèlic 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 "adèlisations" 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 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

$$B^\times\backslash\hat{B}^\times/\hat{R}^\times. \tag{4.2}$$

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 p-adic description of the double coset space appearing in (4.2):

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 $(\dots,1,b_p,1,\dots)$ is a bijection.

For further discussion see ch. III, §4 of [**Vi80**].

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 $\mathcal{H}/\Gamma$ is compact proceeds by constructing a fundamental region for the action of Γ on $\mathcal{H}$ with compact closure, and uses in an essential way the assumption that Γ arises from a quaternion division algebra. It 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 2.10 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 integration 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

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

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

$$T_p(f(z)dz) := \sum_{i=0}^{p} f(\alpha_i^{-1} z) d(\alpha_i^{-1} z). \tag{4.3}$$

- Because the Hecke operators T_n 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 $S_2^{\mathrm{new}}(\Gamma_{N^+,N^-})$ 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* $\mathbb{Q}$; an analogous fact holds for the quotients $\mathcal{H}/\Gamma_{N^+,N^-}$. The way in which this is proved is similar: 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**] or [**Zh01a**], ch. 1).

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 that of theorem 2.9 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 $\mathbb{Q}$. On the other hand the eigenform f determines a homomorphism

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

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*

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

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 a 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 $\mathbb{Q}$ - 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 a quadratic imaginary 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 quadratic imaginary fields. More precisely:

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

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

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 a quadratic imaginary 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 a quadratic imaginary field.

As in chapter 3, given an order $\mathcal{O}$ in a quadratic imaginary 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 a quadratic imaginary field K of discriminant prime to N, and let H/K be the ring class field of K attached to $\mathcal{O}$. Then*

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

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 endomorphism by the maximal order R_0, as well as by $\mathcal{O}$, and these two actions commute with each other. Hence A_τ has endomorphism 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. $\square$

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 a quadratic imaginary 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. $\square$

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 a quadratic imaginary field of discriminant prime to N. If* $\operatorname{sign}(E, K) = -1$, *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 $\operatorname{sign}(E, K) = -1$, the set $S_{E,K}$ has odd cardinality. On the other hand

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

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 $D_n \in \operatorname{Div}^0(CM(\mathcal{O}_n))$ 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. $\square$

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, 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, and $\operatorname{sign}(E, K) = -1$, in constructing the Heegner systems whose existence is predicted by conjecture 3.16?

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 8, relying 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 a quadratic imaginary 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

$$\mathrm{Trace}_{H_n/H_{n'}}(P_\tau)=0. \tag{4.4}$$

   (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 $n=p^t q^s n'$ with $t,s\geq 1$, and that equation (4.4) 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 $\alpha \in B - F$, 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 subalgebra.
(5) Show that the *level* of an Eichler 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 interseciton of two maximal orders R_1 and R_2. (Hint: prove this first for orders in a matrix algebra over a local field.)

CHAPTER 5

Rigid analytic modular forms

The modularity of E has allowed us to construct complex uniformisations

$$\Phi_N : \mathcal{H}/\Gamma_0(N) \longrightarrow E(\mathbb{C}),$$

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

The CM points on these modular and Shimura curves map to a plentiful supply of algebraic points ("Heegner systems") defined over the ring class fields of imaginary quadratic fields.

The purpose of the next two chapters is to enrich this picture further by using the uniformisation Φ_{N^+,N^-} to construct explicit p-adic uniformisations of E by certain discrete arithmetic subgroups of $\mathbf{SL}_2(\mathbb{Q}_p)$, at the primes p dividing N^-.

5.1. p-adic uniformisation

Let p be a prime, let $|\ |_p$ denote the usual normalised p-adic absolute value on $\mathbb{Q}$, and let $\mathbb{Q}_p$ denote the completion of $\mathbb{Q}$ with respect to this absolute value. Choose an algebraic closure $\bar{\mathbb{Q}}_p$ of $\mathbb{Q}_p$. Because $\bar{\mathbb{Q}}_p$ has infinite degree of $\mathbb{Q}_p$, it is no longer complete; the field obtained by completing $\bar{\mathbb{Q}}_p$ with respect to the p-adic distance is a complete algebraically closed field which is commonly denoted $\mathbb{C}_p$ and will play the role of the complex numbers in our analogy.

The *p-adic upper half plane* is defined (as a set) to be

$$\mathcal{H}_p := \mathbb{P}_1(\mathbb{C}_p) - \mathbb{P}_1(\mathbb{Q}_p). \tag{5.1}$$

Note that this set is really analogous to $\mathbb{P}_1(\mathbb{C}) - \mathbb{P}_1(\mathbb{R})$ – two copies of the Poincaré upper half-plane. In the p-adic case this set does not split naturally into two disjoint components, hence it is more appropriate to treat (5.1) as the natural generalisation of the Poincaré upper half plane.

The role of holomorphic functions on $\mathcal{H}$ is played by the so-called *rigid analytic* functions on $\mathcal{H}_p$. These are functions that admit "nice" expressions when restricted to certain distinguished subsets of $\mathcal{H}_p$, called *affinoids*.

We begin by giving prototypical examples of the regions in $\mathcal{H}_p$ (the basic affinoids, and annuli) which are the building blocks for the rigid analytic structure on $\mathcal{H}_p$.

Let

$$\mathrm{red} : \mathbb{P}_1(\mathbb{C}_p) \longrightarrow \mathbb{P}_1(\bar{\mathbb{F}}_p)$$

be the natural map given by reduction modulo the maximal ideal of the ring of integers of $\mathbb{C}_p$. Since $\mathrm{red}(\mathbb{P}_1(\mathbb{Q}_p)) \subset \mathbb{P}_1(\mathbb{F}_p)$, the set

$$\begin{aligned} \mathcal{A} &:= \mathrm{red}^{-1}(\mathbb{P}_1(\bar{\mathbb{F}}_p) - \mathbb{P}_1(\mathbb{F}_p)) \\ &= \{\tau \in \mathcal{H}_p \text{ such that } |\tau - t| \geq 1, \text{ for } t = 0, \ldots, p-1, \text{ and } |\tau| \leq 1\} \end{aligned}$$

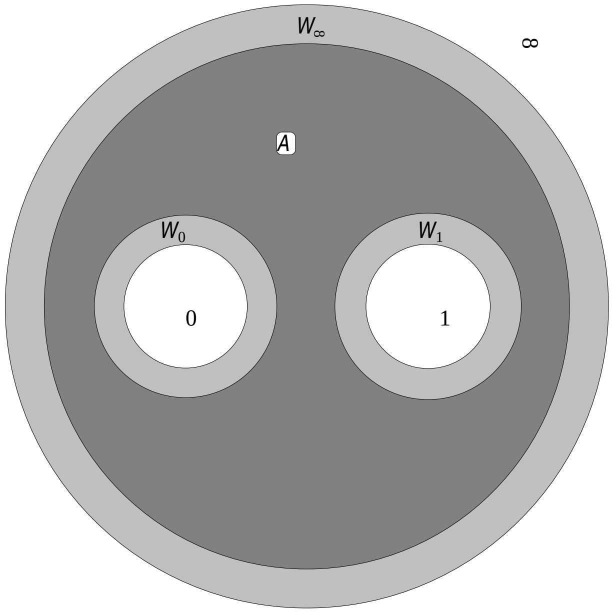

FIGURE 1. The standard affinoid and annuli, when $p = 2$.

is contained in $\mathcal{H}_p$. It is an example of a *standard affinoid* in $\mathcal{H}_p$. It is useful to "thicken" $\mathcal{A}$ by adjoining to it certain *annuli* (which are also subsets of $\mathcal{H}_p$, since the p-adic absolute value on $\mathbb{Q}_p^\times$ takes values in $p^{\mathbb{Z}}$):

$$\begin{aligned} W_t &= \{\tau \text{ such that } \frac{1}{p} < |\tau - t| < 1\}, \quad t = 0, \ldots, p-1, \\ W_\infty &= \{\tau \text{ such that } 1 < |\tau| < p\}. \end{aligned}$$

These regions are illustrated in figure 1, in the most easily drawn case where $p = 2$.

To describe more general affinoids, it is useful to introduce a basic combinatorial structure on $\mathcal{H}_p$: the reduction map.

The target of this map is the so-called *Bruhat-Tits tree* of $\mathbf{PGL}_2(\mathbb{Q}_p)$ denoted $\mathcal{T}$. This is a graph whose vertices are in one-to-one correspondence with similarity classes of $\mathbb{Z}_p$-lattices in $\mathbb{Q}_p^2$. Two vertices are joined by an edge if they can be represented by lattices Λ_1 and Λ_2 satisfying

$$p\Lambda_2 \subset \Lambda_1 \subset \Lambda_2,$$

in which both inclusions are proper. Since this relation is symmetrical it equips $\mathcal{T}$ with the structure of an unordered graph. One can show (cf. exercise 2) that $\mathcal{T}$ is in fact a tree all of whose vertices have valency $p + 1$. This tree is illustrated when $p =$ in the figure below.

The group $\mathbf{PGL}_2(\mathbb{Q}_p)$ acts on $\mathcal{T}$ in the natural way, and this yields an action on $\mathcal{T}$ by graph automorphisms (i.e., isometries with respect to the usual distance function on $\mathcal{T}$).

In our discussion the tree $\mathcal{T}$ is treated as a purely combinatorial object: a collection $\mathcal{T}_0$ of vertices indexed by homothety classes of $\mathbb{Z}_p$-lattices in $\mathbb{Q}_p^2$ and a collection $\mathcal{T}_1$ of edges consisting of pairs of adjacent vertices. An *ordered edge* is an

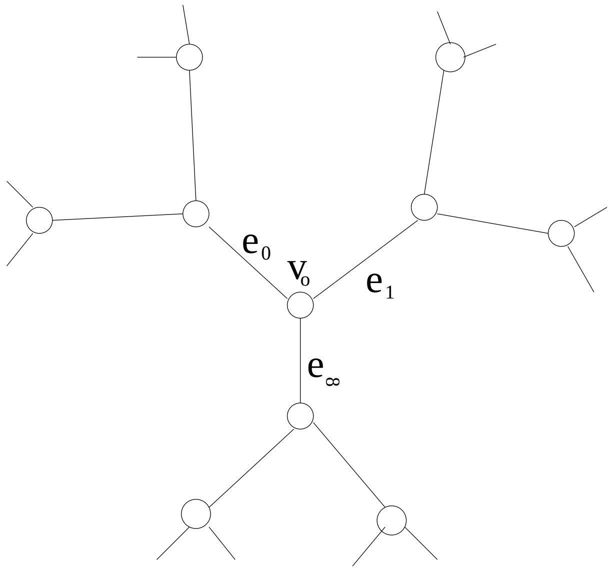

FIGURE 2. The Bruhat-Tits tree of $\mathbf{PGL}_2(\mathbb{Q}_p)$, when $p = 2$

ordered pair $e = (v_1, v_2)$ of adjacent vertices. One then denotes by $s(e) := v_1$ and $t(v) := v_2$ the *source* and *target* of e respectively. Write $\mathcal{E}(\mathcal{T})$ for the set of ordered edges of $\mathcal{T}$.

Let $v_o \in \mathcal{T}_0$ be the distinguished vertex of $\mathcal{T}$ attached to the homothety class of the standard lattice $\mathbb{Z}_p^2 \subset \mathbb{Q}_p^2$. The edges having v_o as endpoint correspond to index p sublattices of $\mathbb{Z}_p^2$ and thus are in canonical bijection with $\mathbb{P}_1(\mathbb{F}_p)$. Label these edges accordingly as $e_0, e_1, \ldots, e_{p-1}, e_\infty \in \mathcal{T}_1$.

PROPOSITION 5.1. *There is a unique map*

$$r : \mathcal{H}_p \longrightarrow \mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1$$

satisfying the following properties:

(1) $r(\tau) = v_o$ *if and only if* $\tau \in \mathcal{A}$;
(2) $r(\tau) = e_t$ *if and only if* $\tau \in W_t$;
(3) r *is* $\mathbf{PGL}_2(\mathbb{Q}_p)$*-equivariant, i.e.,*

$$r(\gamma\tau) = \gamma r(\tau), \qquad \textit{for all } \gamma \in \mathbf{PGL}_2(\mathbb{Q}_p).$$

The proof of this proposition is outlined in exercise 3.

If $e = \{v_1, v_2\}$ is an edge of $\mathcal{T}$, it is convenient to denote by $]e[\subset \mathcal{T}$ the singleton $\{e\}$ and call it the *open edge* attached to e. The subset $[e] := \{e, v_1, v_2\}$ of $\mathcal{T}$ is called the *closed edge* attached to e. Finally, the sets $\mathcal{A}_{[e]} := r^{-1}([e])$ and $W_{]e[} := r^{-1}(]e[)$ are called the *standard affinoid* and the *standard annulus* attached to e respectively. Note that $\mathcal{A}_{[e]}$ is a union of two translates of the standard affinoid $\mathcal{A}$ glued along the annulus $W_{]e[}$. The collection of affinoids $\mathcal{A}_{[e]}$, as e ranges over $\mathcal{T}_1$, gives a covering of $\mathcal{H}_p$ by standard affinoids whose pairwise intersections are either empty or of the form $\mathcal{A}_v := r^{-1}(v)$ with $v \in \mathcal{V}(\mathcal{T})$. The incidence relations in this affinoid covering are thus reflected in the combinatorics of $\mathcal{T}$.

Fix an affinoid $\mathcal{A}_0 \subset \mathcal{H}_p$. A rational function having poles outside $\mathcal{A}_0$ attains its supremum on $\mathcal{A}_0$ (with respect to the p-adic metric). Hence the space of such functions can be equipped with the sup norm.

DEFINITION 5.2. A $\mathbb{C}_p$-valued function f on $\mathcal{H}_p$ is said to be *rigid-analytic* if, for each edge e of $\mathcal{T}$, the restriction of f to the affinoid $\mathcal{A}_{[e]}$ is a uniform limit, with respect to the sup norm, of rational functions on $\mathbb{P}_1(\mathbb{C}_p)$ having poles outside $\mathcal{A}_{[e]}$.

Let Γ be a discrete subgroup of $\mathbf{SL}_2(\mathbb{Q}_p)$. Assume further that the quotient $\mathcal{H}_p/\Gamma$ (with its natural p-adic topology) is compact.

The sheaf of rigid analytic functions on $\mathcal{H}_p/\Gamma$ equips the quotient $\mathcal{H}_p/\Gamma$ with the structure of a *rigid analytic curve* over $\mathbb{Q}_p$, which, by a p-adic analogue of the GAGA theorem, can be identified with an algebraic curve X over $\mathbb{Q}_p$ [**GvdP80**].

REMARK 5.3. Not every curve over $\mathbb{Q}_p$ can be expressed as such a quotient. In fact, it can be shown that if $X = \mathcal{H}_p/\Gamma$ where Γ acts on $\mathcal{T}$ without fixed points, then it has a model over $\mathbb{Z}_p$ whose special fiber is a union of projective lines over $\mathbb{F}_p$ intersecting transversally at ordinary double points. The converse to this statement, due to Mumford, is a p-adic analogue of the classical complex uniformisation theorem:

THEOREM 5.4 (Mumford). *If X is a curve over $\mathbb{Q}_p$ having a model over $\mathbb{Z}_p$ whose special fiber consists of a union of projective lines intersecting at ordinary double points, then there is a discrete group $\Gamma \subset \mathbf{PSL}_2(\mathbb{Q}_p)$ such that the rigid analytic curve $X_{/\mathbb{C}_p}$ is isomorphic to $\mathcal{H}_p/\Gamma$.*

This theorem, which is discussed in [**GvdP80**], will not be used in the sequel, and its statement is included here only for the edification of the reader.

5.2. Rigid analytic modular forms

Let $\Gamma \subset \mathbf{SL}_2(\mathbb{Q}_p)$ be a discrete subgroup as in the previous section.

DEFINITION 5.5. A form of weight k on $\mathcal{H}_p/\Gamma$ is a rigid analytic function f on $\mathcal{H}_p$ such that

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

Denote by $S_k(\Gamma)$ the $\mathbb{C}_p$-vector space of rigid analytic modular forms of weight k with respect to Γ. As in the discussions of chapters 3 and 4, the space $S_2(\Gamma)$ can be identified with the space of rigid analytic differential forms on the quotient $\mathcal{H}_p/\Gamma$, or, equivalently, with the space of regular differential forms on the curve $X_{/\mathbb{C}_p}$. In particular, the dimension of $S_2(\Gamma)$ over $\mathbb{C}_p$ is equal to the genus of this curve.

The classical modular forms in $S_2(\Gamma_0(N))$ studied in chapter 2 could be calculated by exploiting the connection between eigenvalues of Hecke operators and Fourier coefficients of modular forms, together with the action of Hecke operators on the homology of the modular curve made explicit in the theory of modular symbols. There is a method for constructing modular forms in $S_2(\Gamma)$ which, while quite different (there being, for instance, no good notion of Fourier expansion for rigid analytic modular forms) is just as concrete and amenable to calculations as the modular symbol method reviewed in chapter 2. This exploits the connection between forms in $S_2(\Gamma)$ and their associated *p-adic boundary measures*.

More precisely, view $\mathbb{P}_1(\mathbb{Q}_p)$ as the "boundary" of the p-adic upper half plane $\mathcal{H}_p$. This set is endowed with its p-adic topology in which the open balls of the form

$$\begin{aligned} B(a,r) &= \{t \text{ such that } |t-a| < p^{-r}\}, \quad a \in \mathbb{Q}_p, \\ B(\infty,r) &= \{t \text{ such that } |t| > p^r\} \end{aligned}$$

form a basis. These open balls are also compact, and any compact open subset of $\mathbb{P}_1(\mathbb{Q}_p)$ is a finite disjoint union of open balls of the form above.

DEFINITION 5.6. A p-adic distribution on $\mathbb{P}_1(\mathbb{Q}_p)$ is a finitely additive function

$$\mu : \{\text{compact open } U \subset \mathbb{P}_1(\mathbb{Q}_p)\} \longrightarrow \mathbb{C}_p$$

satisfying $\mu(\mathbb{P}_1(\mathbb{Q}_p)) = 0$.

If μ is any p-adic distribution on $\mathbb{P}_1(\mathbb{Q}_p)$, and g is a locally constant function on $\mathbb{P}_1(\mathbb{Q}_p)$, then the integral $\int_{\mathbb{P}_1(\mathbb{Q}_p)} g(t)d\mu(t)$ can be defined in the obvious way as a finite Riemann sum. More precisely, letting

$$\mathbb{P}_1(\mathbb{Q}_p) = U_1 \cup \cdots \cup U_m \tag{5.2}$$

be a decomposition of $\mathbb{P}_1(\mathbb{Q}_p)$ as a disjoint union of open balls such that g is constant on each U_j, one defines

$$\int_{\mathbb{P}_1(\mathbb{Q}_p)} g d\mu := \sum_{j=1}^{m} g(t_j)\mu(U_j), \tag{5.3}$$

where t_j is any sample point in U_j. The distribution relation satisfied by μ ensures that this expression does not depend on the decomposition (5.2) used to define it. It is desirable to impose more stringent regularity properties on a distribution μ so that it can be integrated against a larger class of test functions.

DEFINITION 5.7. A p-adic *measure* is a bounded distribution, i.e., a distribution for which there is a constant C satisfying

$$|\mu(U)|_p < C, \text{ for all compact open } U \subset \mathbb{P}_1(\mathbb{Q}_p).$$

If λ is any continuous function on $\mathbb{P}_1(\mathbb{Q}_p)$, then the integral of λ against μ can be defined by the rule

$$\int_{\mathbb{P}_1(\mathbb{Q}_p)} \lambda(t)d\mu(t) = \lim_{\mathcal{C}=\{U_\alpha\}} \sum_\alpha \lambda(t_\alpha)\mu(U_\alpha), \tag{5.4}$$

where the limit is taken over increasingly fine covers $\{U_\alpha\}$ of $\mathbb{P}_1(\mathbb{Q}_p)$ by disjoint compact open subsets U_α, and t_α is a sample point in U_α.

The connection between boundary measures and rigid analytic functions on $\mathcal{H}_p$ is given by the following lemma:

LEMMA 5.8. (1) *If μ is a measure on $\mathbb{P}_1(\mathbb{Q}_p)$, then*

$$f_\mu(z) = \int_{\mathbb{P}_1(\mathbb{Q}_p)} \left(\frac{1}{z-t}\right) d\mu(t)$$

is a rigid analytic function on $\mathcal{H}_p$.

(2) *If μ is a Γ-invariant measure on $\mathbb{P}_1(\mathbb{Q}_p)$, then f_μ belongs to $S_2(\Gamma)$.*

Proof. Part 1 follows directly from the definition of a rigid analytic function on $\mathcal{H}_p$, since the integral defining $f_\mu(z)$ is expressed through (5.4) as a limit of rational functions having poles in $\mathbb{P}_1(\mathbb{Q}_p)$, and since the distribution and boundedness satisfied by μ ensures that this convergence is uniform on each affinoid in $\mathcal{H}_p$. To show part 2, note that both $\frac{1}{\gamma z - t}$ and $\frac{1}{z-\gamma^{-1}t}$ are rational functions of t having a unique simple pole at $t = \gamma z$, the first with residue (-1) and the second with residue $-(cz+d)^{-2}$. Hence

$$\frac{1}{\gamma z - t} - (cz+d)^2\left(\frac{1}{z-\gamma^{-1}t}\right) = C,$$

where C is a constant (possibly depending on z and γ but not on t). It follows that for any $\gamma \in \Gamma$:

$$\begin{aligned} f_\mu(\gamma z) &= \int_{\mathbb{P}_1(\mathbb{Q}_p)} \left(\frac{1}{\gamma z - t}\right) d\mu(t) \\ &= (cz+d)^2 \int_{\mathbb{P}_1(\mathbb{Q}_p)} \left(\frac{1}{z-\gamma^{-1}t}\right) d\mu(t) = (cz+d)^2 f_\mu(z). \end{aligned}$$

The second equality uses the fact that the total measure of $\mathbb{P}_1(\mathbb{Q}_p)$ is 0, and the third exploits the invariance of μ under Γ. □

Denote by $\mathrm{Meas}(\mathbb{P}_1(\mathbb{Q}_p), \mathbb{C}_p)^\Gamma$ the space of all Γ-invariant measures on $\mathbb{P}_1(\mathbb{Q}_p)$.

Theorem 5.9 (Schneider, Teitelbaum). *The assignment $\mu \mapsto f_\mu$ is an isomorphism from* $\mathrm{Meas}(\mathbb{P}_1(\mathbb{Q}_p), \mathbb{C}_p)^\Gamma$ *to* $S_2(\Gamma)$.

Sketch of Proof. The construction of the inverse map can in fact be made explicit. To do this, note that an ordered edge e of $\mathcal{T}$ determines a subtree $\mathcal{T}_e$ of $\mathcal{T}$, defined to be the largest connected subtree containing e and no other edge having $s(e)$ as an endpoint. Let

$$\Sigma_e := r^{-1}(\mathcal{T}_e) \subset \mathcal{H}_p,$$

and let $\bar{\Sigma}_e$ denote the closure of Σ_e in $\mathbb{P}_1(\mathbb{C}_p)$. Finally let $U_e := \bar{\Sigma}_e \cap \mathbb{P}_1(\mathbb{Q}_p)$. The assignment $e \mapsto U_e$ sets up a correspondence between ordered edges of $\mathcal{T}$ and compact open balls in $\mathbb{P}_1(\mathbb{Q}_p)$. Given a rigid analytic function f on $\mathcal{H}_p$, define a distribution μ_f on $\mathbb{P}_1(\mathbb{Q}_p)$ by the rule

$$\mu_f(U_e) = \lim_{j \longrightarrow \infty} \sum_{x \in \bar{\Sigma}_e} \mathrm{res}_x(f_j(z)dz),$$

where f_j is a sequence of rational functions with poles outside $\mathcal{A}_{s(e)}$ which converges uniformly to f on $\mathcal{A}_{s(e)}$. The residue theorem implies that μ_f satisfies a distribution relation, and the Γ-invariance of $f(z)dz$ implies the corresponding Γ-invariance of μ_f. In effect, the *boundary measure* μ_f attached to f encodes the p-adic residues of f. The verification that $f_{\mu_f} = f$ is left to the reader as an exercise. (Cf. exercise 5.) □

The construction above leads to a useful description of $\mathrm{Meas}(\mathbb{P}_1(\mathbb{Q}_p), \mathbb{C}_p)^\Gamma$ which has the virtue of laying bare the simple combinatorial nature of this object.

Definition 5.10. A *harmonic cocycle* on $\mathcal{T}$ is a function $c : \mathcal{E}(\mathcal{T}) \longrightarrow \mathbb{C}_p$ satisfying

(1) $c(e) = -c(\bar{e})$, for all $e \in \mathcal{E}(\mathcal{T})$;
(2) $\sum_{s(e)=v} c(e) = 0; \quad \sum_{t(e)=v} c(e) = 0$, for all $v \in \mathcal{T}_0$.

(Note that the second part of condition (2) is redundant since it follows from (1) and the first part of (2).)

A harmonic cocycle c gives rise to a distribution μ_c on $\mathbb{P}_1(\mathbb{Q}_p)$ by the rule

$$\mu_c(U_e) = c(e). \tag{5.5}$$

Conversely, c can be recovered from the associated distribution by the rule above. Under this bijection the Γ-invariant distributions correspond to Γ-invariant harmonic cocycles on $\mathcal{T}$.

The finite-dimensionality of $\mathrm{Meas}(\mathbb{P}_1(\mathbb{Q}_p), \mathbb{C}_p)^\Gamma$ (and hence, of $S_2(\Gamma)$) can now be seen to follow directly from the following lemma:

LEMMA 5.11. *The quotient $\mathcal{T}/\Gamma$ is a finite graph.*

PROOF. See exercise 6. $\square$

The space $\mathrm{Meas}(\mathbb{P}_1(\mathbb{Q}_p), \mathbb{Z})^\Gamma$ of Γ-invariant $\mathbb{Z}$-valued distributions on $\mathbb{P}_1(\mathbb{Q}_p)$, corresponding to $\mathbb{Z}$-valued harmonic cocycles, yields a natural integral structure on $\mathrm{Meas}(\mathbb{P}_1(\mathbb{Q}_p), \mathbb{C}_p)^\Gamma$. Its image in $S_2(\Gamma)$, denoted by $S_2(\Gamma)^{\mathbb{Z}}$, plays a role somewhat similar to that of modular forms with integral Fourier coefficients in the theory of modular forms on $\Gamma_0(N)$.

5.3. p-adic line integrals

Let f be a rigid analytic function on $\mathcal{H}_p$. The goal of this section is to define a good notion of p-adic line integral attached to such an object. This line integral should be an expression of the form $\int_{\tau_1}^{\tau_2} f(z)dz \in \mathbb{C}_p$ obeying the same formal properties of the complex line integral, namely it should be linear in f and additive in the endpoints of integration:

$$\int_{\tau_1}^{\tau_2} f(z)dz + \int_{\tau_2}^{\tau_3} f(z)dz = \int_{\tau_1}^{\tau_3} f(z)dz, \quad \forall \tau_1, \tau_2, \tau_3 \in \mathcal{H}_p.$$

If $f(z)dz = dF$ is an exact differential on $\mathcal{H}$, one would clearly like to define

$$\int_{\tau_1}^{\tau_2} f(z)dz = F(\tau_2) - F(\tau_1). \tag{5.6}$$

The equation $dF = f(z)dz$ is sufficient to define F up to a locally constant, hence constant, function in the complex setting. A difficulty arises from the circumstance that in the p-adic context, there are plenty of locally constant functions which are not constant, because $\mathcal{H}_p$ is totally disconnected. This leads to an ambiguity in the choice of F.

One may try to remedy this situation by singling out a particularly natural antiderivative of certain functions. For example, among all possible choices of function F on $\mathbb{C}_p$ such that $dF = \frac{dz}{z}$, the p-adic logarithm defined in the open disc in $\mathbb{C}_p$ of radius 1 centered at 1 by the power series

$$\log(1-z) = \sum_{n=1}^{\infty} \frac{z^n}{n}$$

is singled out by the property of being a homomorphism from this open disc (under multiplication) to $\mathbb{C}_p$. It is customary to choose an extension of the p-adic logarithm to all of $\mathbb{C}_p$ (a "branch")

$$\log : \mathbb{C}_p^\times \longrightarrow \mathbb{C}_p$$

by fixing some element $\pi \in \mathbb{C}_p^\times$ satisfying $|\pi|_p < 1$ and decreeing that $\log(\pi) = 0$ (and requiring, of course, that log be a homomorphism on $\mathbb{C}_p^\times$). The standard choice is obtained by taking $\pi = p$, but this may not always be the most natural choice in all situations.

Having fixed a choice of p-adic logarithm, one has, for each rational differential $f(z)dz$ on $\mathbb{P}_1(\mathbb{C}_p)$, a formal antiderivative of the form

$$F(z) = R(z) + \sum_{j=1}^{t} \lambda_j \log(z - z_j),$$

where R is a rational function, the λ_j belong to $\mathbb{C}_p$, and the z_j are the poles of $f(z)dz$. This antiderivative is unique up to an additive constant, and hence equation (5.6) can be used to write down a well-defined line integral attached to $f(z)dz$. Extending this definition by continuity to all rigid analytic functions leads to the following definition.

Definition 5.12. Let f be a rigid analytic function on $\mathcal{H}_p$. Assume that its associated boundary distribution μ_f is a measure. Then the *p-adic line integral* attached to $f(z)dz$ is defined to be

$$\int_{\tau_1}^{\tau_2} f(z)dz := \int_{\mathbb{P}_1(\mathbb{Q}_p)} \log\left(\frac{t-\tau_2}{t-\tau_1}\right) d\mu_f(t). \tag{5.7}$$

Rephrasing the dicussion which precedes definition 5.12, one can seek to justify this definition *a posteriori* through the following formal computation:

$$\int_{\tau_1}^{\tau_2} f(z)dz = \int_{\tau_1}^{\tau_2} \int_{\mathbb{P}_1(\mathbb{Q}_p)} \left(\frac{dz}{z-t}\right) d\mu_f(t) = \int_{\mathbb{P}_1(\mathbb{Q}_p)} \log\left(\frac{t-\tau_2}{t-\tau_1}\right) d\mu_f(t).$$

Note that this definition depends crucially on the choice of the branch of the p-adic logarithm that was made in fixing a primitive for dz/z.

In the special case where f is attached to a $\mathbb{Z}$-valued distribution, i.e., where f belongs to $S_2(\Gamma)^{\mathbb{Z}}$, it can be useful to "work multiplicatively" by formally exponentiating (5.7) above and setting:

$$⨰_{\tau_1}^{\tau_2} f(z)dz = ⨰_{\mathbb{P}_1(\mathbb{Q}_p)} \left(\frac{t-\tau_2}{t-\tau_1}\right) d\mu_f(t) := \lim_{\mathcal{C}=\{U_\alpha\}} \prod_\alpha \left(\frac{t_\alpha-\tau_2}{t_\alpha-\tau_1}\right)^{\mu_f(U_\alpha)}, \tag{5.8}$$

where the limit in the last expression is taken over increasingly fine covers $\mathcal{C} = \{U_\alpha\}$ of $\mathbb{P}_1(\mathbb{Q}_p)$ by disjoint compact open subsets, and the t_α are sample points in U_α. The multiplicative integral is related to its additive counterpart by the rule

$$\int_{\tau_1}^{\tau_2} f(z)dz = \log\left(⨰_{\tau_1}^{\tau_2} f(z)dz\right),$$

but it carries more information, as the p-adic logarithm is not injective. Also, it is more canonical than its additive counterpart since it does not depend on a choice of a branch of the p-adic logarithm. However, its definition relies crucially on the integrality of the boundary distribution atttached to f.

The multiplicative integral can be used to define a p-adic analogue of the classical Abel-Jacobi map

$$\Phi_{AJ} : \mathrm{Div}^0(\mathcal{H}_p) \longrightarrow \hom(S_2(\Gamma)^{\mathbb{Z}}, \mathbb{C}_p^\times) \simeq (\mathbb{C}_p^\times)^g,$$

following [**GvdP80**]. The map Φ_{AJ} is defined by sending a divisor D of degree 0 to the functional

$$\omega \mapsto \times\!\!\!\!\!\!\int_D \omega$$

which belongs to $\hom(S_2(\Gamma)^{\mathbb{Z}}, \mathbb{C}_p^\times)$. It can be shown that Φ_{AJ} maps the group of divisors on $\mathcal{H}_p$ which become 0 on $\mathcal{H}_p/\Gamma$ to a lattice Λ in this group, so that it gives rise, by passing to the quotient, to an analytic map of abelian varieties over $\mathbb{C}_p$:

$$\Phi_{AJ} : \mathrm{Jac}(\mathcal{H}_p/\Gamma) \longrightarrow \hom(S_2(\Gamma)^{\mathbb{Z}}, \mathbb{C}_p^\times)/\Lambda.$$

In the next chapter, we will identify certain special arithmetic $\Gamma \subset \mathbf{SL}_2(\mathbb{Q}_p)$ arising from p-units in orders in definite quaternion algebras, and use them to obtain p-adic Weil uniformisations

$$\mathrm{Div}^0(\mathcal{H}_p/\Gamma) \longrightarrow E(\mathbb{C}_p)$$

attached to (modular) elliptic curves E over $\mathbb{Q}$.

Further results

The book of Gerritzen and van der Put [**GvdP80**] provides a good introduction p-adic Schottky groups and to Mumford's theory of p-adic uniformisation. An account of the p-adic Poisson kernel, as well as an extension of the theory presented here to modular forms of higher weight, can be found in the articles [**Te90**] and [**Sch84**].

Exercises

(1) Show that $\bar{\mathbb{Q}}_p$ is not complete in the p-adic topology by exhibiting a Cauchy sequence in this field which does not converge.

(2) Show that the graph $\mathcal{T}$ is a tree all of whose vertices have valency equal to $p+1$.

(3) Let $G = \mathbf{PGL}_2(\mathbb{Q}_p)$ act by Möbius transformations on $\mathcal{H}_p$, and by left translations on $\mathcal{T}$. Let $G_0 = \mathbf{PGL}_2(\mathbb{Z}_p)$ be the maximal compact subgroup of G and let G_1 be the subgroup of G_0 represented by matrices which are upper-triangular modulo p.
   (a) Show that G_0 is precisely the group of elements in G which preserve the standard afinoid $\mathcal{A} \subset \mathcal{H}_p$, and that G_1 is the stabiliser of the annulus W_∞ in G_0.
   (b) Show that G_0 (resp. G_1) is the stabiliser in G of the standard vertex v_o (resp. of the ordered edge e_∞).
   (c) Prove proposition 5.1.

(4) Check that the expression in (5.4) converges in $\mathbb{C}_p$.

(5) Complete the proof of theorem 5.9 by showing that $f_{\mu_f} = f$ where μ_f is the boundary distribution obtained in the proof of theorem 5.9 by taking the residues of f.

(6) Endow $\mathcal{H}_p$ with the p-adic topology, and $\mathcal{T}$ with the topology in which a subset U is said to be open if, for any vertex $v \in U$, the edges having v as endpoint also belong to U. Show that the reduction map $r : \mathcal{H}_p \longrightarrow \mathcal{T}$ is continuous for these topologies. Conclude that the quotient graph $\mathcal{T}/\Gamma$ is a finite graph if $\mathcal{H}_p/\Gamma$ is compact.

CHAPTER 6

Rigid analytic modular parametrisations

The parametrisations alluded to in the title of this chapter arise from the fact that, for the primes p which divide N^-, the Shimura curves X_{N^+,N^-} introduced in chapter 4 are equipped with p-adic uniformisations discovered by Čerednik and Drinfeld which ressemble the complex analytic descriptions that were already given. These uniformisations can be used to construct p-adic modular parametrisations for elliptic curves over $\mathbb{Q}$ whose conductors satisfy suitable conditions.

6.1. Rigid analytic modular forms on quaternion algebras

Let N be a positive integer and p a prime number dividing N exactly.

DEFINITION 6.1. A p-admissible factorisation of N is a factorisation of the form $N = pN^+N^-$, where

(1) The integers p, N^+, and N^- are pairwise coprime;
(2) The integer N^- is square-free and the product of an odd number of primes.

To each p-admissible factorisation $N = pN^+N^-$ is attached a subgroup $\Gamma^{(p)}_{N^+,N^-}$ of $\mathbf{SL}_2(\mathbb{Q}_p)$ as follows. Let B be the *definite* quaternion algebra ramified precisely at the primes dividing N^-, together with the archimedean place. Such a quaternion algebra exists by proposition 4.1. It follows from proposition 4.4 that there is a unique Eichler $\mathbb{Z}[1/p]$-order R of level N^+ in B, up to conjugation by $B^\times$. Since B is split at the prime p, it is possible to choose an identification

$$\iota : B \otimes \mathbb{Q}_p \longrightarrow M_2(\mathbb{Q}_p).$$

Letting $R_1^\times$ denote the group of elements in R of reduced norm 1, the group $\Gamma^{(p)}_{N^+,N^-}$ is then defined to be

$$\Gamma^{(p)}_{N^+,N^-} := \iota(R_1^\times) \subset \mathbf{SL}_2(\mathbb{Q}_p).$$

It can be proved that $\Gamma^{(p)}_{N^+,N^-}$ acts on $\mathcal{H}_p$ with compact quotient so that the theory of rigid analytic modular forms developed in chapter 5 applies to forms on this group. In particular the space $S_2(\Gamma^{(p)}_{N^+,N^-})^{\mathbb{Z}}$ is a finitely generated $\mathbb{Z}$-module of rank equal to the genus of the rigid analytic curve $\mathcal{H}_p/\Gamma^{(p)}_{N^+,N^-}$.

The space $S_2(\Gamma^{(p)}_{N^+,N^-})$ is in addition endowed with the action of Hecke operators T_ℓ which are defined as before in terms of double coset decompositions. More precisely, for each prime ℓ which does not divide N, choose an element $\alpha_\ell \in R$ of reduced norm ℓ, write the double coset $\Gamma^{(p)}_{N^+,N^-}\alpha_\ell\Gamma^{(p)}_{N^+,N^-}$ as a disjoint union of left cosets

$$\Gamma^{(p)}_{N^+,N^-}\alpha_\ell\Gamma^{(p)}_{N^+,N^-} = \cup_{i=1}^{\ell+1}\gamma_i\Gamma^{(p)}_{N^+,N^-},$$

and set

$$T_\ell f(z)dz := \sum_{i=1}^{\ell+1} f(\gamma_i^{-1}z)d(\gamma_i^{-1}z).$$

DEFINITION 6.2. An eigenform f is said to be *normalised* if it belongs to $S_2(\Gamma_{N^+,N^-}^{(p)})^{\mathbb{Z}}$ and is not divisible by any integer > 1 in this group.

Note that, if f is normalised then so is $-f$ and there seems to be no natural way of resolving this ambiguity in signs. Note also that this normalisation applies only to eigenforms with rational residues, whose asociated Hecke eigenvalues necessarily belong to $\mathbb{Z}$.

6.2. The Čerednik-Drinfeld theorem

The rigid analytic curve $\mathcal{H}_p/\Gamma_{N^+,N^-}^{(p)}$ shares some common features with the quotient $\mathcal{H}/\Gamma_{N^+,N^-}$ that arose in decribing the complex uniformisation of the Shimura curve X_{N^+,N^-p}, most notably the presence of a large ring of correspondences given by Hecke operators indexed by the primes ℓ that do not divide N. The following theorem of Čerednik and Drinfeld reveals that this analogy runs deeper:

THEOREM 6.3. *The rigid analytic quotient $\mathcal{H}_p/\Gamma_{N^+,N^-}^{(p)}$ is isomorphic (as an algebraic curve over $\mathbb{C}_p$) to X_{N^+,N^-p}.*

REMARK 6.4. More precisely, the quotient $\mathcal{H}_p/\Gamma_{N^+,N^-}^{(p)}$ is identified with the $\mathbb{C}_p$-points of an algebraic curve X defined over $\mathbb{Q}_p$. The curve X becomes isomorphic to X_{N^+,N^-p} over the unramified quadratic extension of $\mathbb{Q}_p$.

Theorem 6.3 asserts that the quotients $\mathcal{H}_p/\Gamma_{N^+,N^-}^{(p)}$ and $\mathcal{H}/\Gamma_{N^+,N^-p}$ decribe – over $\mathbb{C}_p$ and $\mathbb{C}$ respectively – the *same* algebraic curve, even though the groups $\Gamma_{N^+,N^-}^{(p)}$ and Γ_{N^+,N^-p} are defined in terms of different quaternion algebras. It was first proved by Čerednik [**Ce76**] building on ideas of Ihara [**Ih68**], (the same which partly inspired the point of view adopted in chapter 9; cf also [**Ih79**]). A more conceptual proof relying on an interpretation of $\mathcal{H}_p$ as classifying certain formal groups was later given by Drinfeld [**Dr76**]. The standard reference for the Čerednik-Drinfeld theorem, and particularly for Drinfeld's approach, is the exposition given in [**BC92**].

6.3. The p-adic Shimura-Taniyama-Weil conjecture

Let E be an elliptic curve of conductor N and let pN^+N^- be a p-admissible factorisation of N. Let $\Gamma_{N^+,N^-}^{(p)} \subset \mathbf{SL}_2(\mathbb{Q}_p)$ be the discrete subgroup arising from this factorisation.

THEOREM 6.5 ("Rigid Shimura-Taniyama-Weil "). *There exists a unique (up to sign) normalised eigenform $f \in S_2(\Gamma_{N^+,N^-}^{(p)})^{\mathbb{Z}}$ such that*

$$T_\ell(f) = a_\ell(E)f,$$

for all $\ell \not| N$.

SKETCH OF PROOF. By the Shimura-Taniyama-Weil theorem in the case of Shimura curves (theorem 4.13 of chapter 4) there is an eigenform g on $\mathcal{H}/\Gamma_{N^+,N^-p}$ attached to E, which is unique up to multiplication by $\mathbb{C}^\times$; but theorem 6.3 implies that the Hecke modules $S_2(\Gamma_{N^+,N^-}^{(p)})$ and $S_2(\Gamma_{N^+,N^-p})$ give rise to the same systems of eigenvalues for the Hecke operators, and with the same multiplicities, since they are both identified with the space of regular differentials on the curve X_{N^+,N^-p} (over $\mathbb{C}_p$ and $\mathbb{C}$ respectively). Therefore there is associated to g a Hecke eigenform $f \in S_2(\Gamma^{(p)}_{N^+,N^-})$ which is unique up to multiplication by an element of $\mathbb{C}_p^\times$. Since the Hecke eigenvalues are integers, f can be rescaled to be a normalised eigenform in $S_2(\Gamma^{(p)}_{N^+,N^-})^{\mathbb{Z}}$, and it is then uniquely determined up to sign. □

REMARK 6.6. It is not strictly necessary to invoke the Čerednik-Drinfeld theorem in this proof. In fact, the Jacquet-Langlands correspondence – which in the case at hand follows from an earlier result of Eichler – makes it possible to establish directly the existence of a $\Gamma^{(p)}_{N^+,N^-}$-invariant measure μ_f (or, equivalently, of the associated harmonic cocycle c_f) attached to f, given the knowledge that E corresponds to a classical modular form on $\mathcal{H}/\Gamma_0(N)$.

Theorem 6.5 implies the existence of a *rigid analytic Weil uniformisation*

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

which is defined as follows. Firstly, the function

$$\mathrm{Div}^0(\mathcal{H}_p) \longrightarrow \mathbb{C}_p^\times, \qquad D \mapsto \times\!\!\!\!\!\!\int_D f(z)dz$$

maps the group of divisors which become trivial in $\mathcal{H}_p/\Gamma^{(p)}_{N^+,N^-}$ to a lattice in $\mathbb{C}_p^\times$ generated by an element $q \in \mathbb{Q}_p^\times$. The Tate curve $E_q := \mathbb{C}_p^\times/q^{\mathbb{Z}}$ is related to E by an isogeny β which is defined over $\mathbb{C}_p$, and letting

$$\Phi_{\mathrm{Tate}} : \mathbb{C}_p^\times \longrightarrow E_q(\mathbb{C}_p)$$

be the Tate uniformisation one sets

$$\Phi^{(p)}_{N^+,N^-}(D) := \beta\left(\Phi_{\mathrm{Tate}}\left(\times\!\!\!\!\!\!\int_D f(z)dz\right)\right). \tag{6.1}$$

To actually calculate this map only requires a knowledge of the residues attached to $f(z)dz$ (and of the isogeny β). Section 6.5 illustrates through an example how this can sometimes be achieved in practice.

6.4. Complex multiplication, revisited

The notion of CM points on $\mathcal{H}_p/\Gamma^{(p)}_{N^+,N^-}$, and a p-adic analytic Heegner point construction, can be formulated for the rigid analytic modular parametrisation $\Phi^{(p)}_{N^+,N^-}$ in a way which is pleasingly similar to the complex setting described in chapter 4.

DEFINITION 6.7. The *associated order* of $\tau \in \mathcal{H}_p$ is the set

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

As before, one can show that $\mathcal{O}_\tau$ is isomorphic either to $\mathbb{Z}[1/p]$ or to a $\mathbb{Z}[1/p]$-order in a quadratic subfield K of B. Also, $\mathcal{O}_\tau$ is equipped with a canonical inclusion into $\mathbb{C}_p$ which to γ associates the eigenvalue of $\iota(\gamma)$ acting on the column vector $(\tau, 1)$.

REMARK 6.8. Since B is a definite quaternion algebra, note that the field K is necessarily a quadratic imaginary field.

DEFINITION 6.9. A point $\tau \in \mathcal{H}_p/\Gamma^{(p)}_{N^+,N^-}$ is called a CM point if its associated order is a $\mathbb{Z}[1/p]$-order in a quadratic imaginary field.

Fixing such an order $\mathcal{O}$, we write

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

LEMMA 6.10. *Let $\mathcal{O}$ be an order of discriminant prime to N, with fraction field K. Then $CM(\mathcal{O})$ is non empty if and only if*

(1) *K is a quadratic imaginary extension of $\mathbb{Q}$;*
(2) *all the primes dividing N^-p are inert in K;*
(3) *all the primes dividing N^+ are split in K.*

PROOF. The proof is almost identical to that of lemma 4.17 of chapter 4, except that the roles of the places p and ∞ are interchanged. The reader may find it instructive to fill in the details of the proof. (See exercise 5.) □

Since p is inert in K, the Picard group of rank one projective modules over $\mathcal{O}$ is equal to the Picard group of the order $\mathcal{O} \cap \mathcal{O}_K$ of K. Let H be the ring class field of K attached to $\mathcal{O}$. Note that its conductor is prime to p. Hence p (viewed as a prime of K) *splits completely* in H/K.

The following can be viewed as a p-adic variant of the main theorem of complex multiplication for Shimura curves as formulated in theorem 4.16 of chapter 4.

THEOREM 6.11. *Let $\mathcal{O}$ be a $\mathbb{Z}[1/p]$-order in a quadratic imaginary field K, of discriminant prime to N, and let H/K be the associated ring class field, viewed as a subfield of $\mathbb{C}_p$. Then*

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

IDEA OF PROOF. Under the moduli interpretation of X_{N^+,N^-p} given by Drinfeld's theory, points $\tau \in CM(\mathcal{O})$ correspond to points in X_{N^+,N^-p} which are moduli of abelian surfaces with endomorphism by $M_2(\mathcal{O}_0)$. These surfaces are isomorphic to a product $A \times A$ of an elliptic curve A with CM by $\mathcal{O}_0$, with itself. Given this fact, the result follows from the usual theory of complex multiplication. Further details are explained in [**BD98**]. □

6.5. An example

We now describe an example to illustrate how the rigid analytic modular parametrisations can be computed and used to find algebraic points on elliptic curves, in practice.
Let E be the elliptic curve of conductor $N = 14$ given by the minimal Weierstrass equation

$$y^2 + xy + y = x^3 + 4x - 6. \tag{6.2}$$

The triple $(p, N^+, N^-) = (7, 1, 2)$ is a 7-admissible factorisation of 14.

Let B denote the algebra of rational Hamilton quaternions

$$B = \mathbb{Q} + \mathbb{Q}i + \mathbb{Q}j + \mathbb{Q}k,$$

a quaternion algebra over $\mathbb{Q}$ which is ramified precisely at 2 and ∞. Let R_0 be Hurwitz's ring of integral quaternions:

$$R_0 = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}\omega,$$

where $\omega = \frac{1+i+j+k}{2}$. The main facts about R_0 that will be needed are

THEOREM 6.12 (Hurwitz). *The ring R_0 is the unique maximal order in B up to conjugation in $B^\times$. Every left R_0-ideal is principal.*

PROOF. See for example [**Gr87**], §1 and 2. □

In particular, the ring $R := R_0[1/7]$ is the unique maximal $\mathbb{Z}[1/7]$-order in B up to conjugation in $B^\times$. Fix an isomorphism ι of $B_7 := B \otimes \mathbb{Q}_7$ with $M_2(\mathbb{Q}_7)$ which has the property that $\iota^{-1}(M_2(\mathbb{Z}_7)) = R_0 \otimes \mathbb{Z}_7$. To fix ideas, we may take

$$\iota(i) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \iota(j) = \begin{pmatrix} \rho & \rho+1 \\ \rho+1 & -\rho \end{pmatrix}, \quad \iota(k) = \begin{pmatrix} \rho+1 & -\rho \\ -\rho & -\rho-1 \end{pmatrix},$$

where $\rho = \lim_{n\longrightarrow\infty} 2^{7^n}$ is a primitive cube root of unity in $\mathbb{Z}_7$.

Let $\mathcal{T}_7 = \mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1$ denote the Bruhat-Tits tree of $\mathbf{PGL}_2(\mathbb{Q}_7)$ introduced in chapter 5, and let $\Gamma = \iota(R_1^\times) \subset \mathbf{PSL}_2(\mathbb{Q}_7)$. This group acts naturally on $\mathcal{T}$; the following lemma yields a precise description of the quotient graph $\mathcal{T}/\Gamma$.

LEMMA 6.13. *The group Γ has precisely two orbits acting on $\mathcal{T}_0$. Likewise the set $\Gamma\backslash\mathcal{T}_1$ also has cardinality two.*

PROOF. Recall the distinguished vertex v_o attached to the standard lattice $\mathbb{Z}_7^2$. The map $\gamma \mapsto \gamma v_0$ identifies the coset space $\mathbf{PGL}_2(\mathbb{Q}_7)/\mathbf{PGL}_2(\mathbb{Z}_7)$ with $\mathcal{T}_0$. Hence $\Gamma\backslash\mathcal{T}_0$ can be identified with

$$\iota(R_1^\times)\backslash\mathbf{PGL}_2(\mathbb{Q}_7)/\mathbf{PGL}_2(\mathbb{Z}_7) = R_1^\times\backslash B_7^\times/R_{0,7}^\times\mathbb{Q}_7^\times, \tag{6.3}$$

where as before $R_{0,7} := R_0 \otimes \mathbb{Z}_7$ and $R_1^\times$ denotes the group of elements of reduced norm 1 in R. Since $R_1^\times\backslash R^\times/R_{0,7}^\times\mathbb{Q}_7^\times$ has cardinality two – with cosets consisting of elements of $R^\times$ whose determinant is a unit of $\mathbb{Z}_7$ multiplied by an even (odd) power of 7 – the set on the right of (6.3) admits a natural two-to-one map to the coset space

$$R^\times\backslash B_7^\times/R_{0,7}^\times\mathbb{Q}_7^\times.$$

Noting that $R = R_0[1/7]$ and that the prime 7 is split in the quaternion algebra B, theorem 4.5 implies that this coset space is equal to

$$B^\times\backslash\hat{B}^\times/\hat{R}_0^\times. \tag{6.4}$$

But by the discussion preceeding the statement of theorem 4.5 of chapter 4, the double coset space of (6.4) is in bijection with the set of conjugacy classes of maximal orders in B. This set has cardinality one by theorem 6.12 above. In fact, the quotient $\Gamma\backslash\mathcal{T}_0$ admits a simple description: one orbit consists of vertices which are at an even distance from the vertex v_o, and the second orbit consists of vertices which are at an odd distance from v_0. To study $\Gamma\backslash\mathcal{T}_1$, note that any edge of $\mathcal{T}$ is equivalent under Γ to an edge having v_o as endpoint, i.e., to one of the edges $e_0, \ldots e_6, e_\infty$ naturally indexed by $\mathbb{P}_1(\mathbb{F}_7)$ that were introduced in chapter 5. Since the stabiliser of v_0 in Γ is equal to $\iota(R_0^\times)$, it follows that $\Gamma \subset \mathcal{T}_1$ is naturally identified with the orbit space

$$\iota(R_0^\times)\backslash\mathbb{P}_1(\mathbb{F}_7).$$

The group $R_0^\times$ is a group of order 24, generated by i, j, and ω, and $\iota(R_0^\times)$ acts on $\mathbb{P}_1(\mathbb{F}_7)$ in the obvious way. This action breaks $\mathbb{P}_1(\mathbb{F}_7)$ into two orbits of the same cardinality, $\{\infty, 0, 2, 3\}$ and $\{1, 4, 5, 6\}$. Hence $\Gamma\backslash\mathcal{T}_1$ has two orbits, E_1 and E_2, defined by

$$e \in \begin{cases} E_1 & \text{if} \quad e \text{ is } \Gamma\text{-equivalent to } e_0, e_2, e_3, \text{ or } e_\infty. \\ E_2 & \text{if} \quad e \text{ is } \Gamma\text{-equivalent to } e_1, e_4, e_5, \text{ or } e_6. \end{cases}$$

□

Lemma 6.13 allows the determination of the full set of Γ-invariant harmonic cocycles on $\mathcal{T}$. Firstly, such a cocycle c is completely determined by its values on the ordered edges $e'_0, \ldots, e'_\infty$ having the same endpoints as $e_0, \ldots, e_\infty$ respectively, and ordered by setting $s(e'_j) = v_o$. Secondly, the harmonicity and Γ-invariance conditions force the relations

$$c(e'_0) = c(e'_2) = c(e'_3) = c(e'_\infty) = -c(e'_1) = -c(e'_4) = -c(e'_5) = -c(e'_6).$$

Conversely, there is a non-zero Γ-invariant cocycle satisfying the relations above. Normalise it so that $c(e'_0) = 1$. The $\mathbb{Z}$-module $S_2(\Gamma)^{\mathbb{Z}}$ has rank one, and is generated by the rigid analytic modular form f whose residues are given by the harmonic cocycle c.

The first few Hecke operators acting on c can be evaluated explicitly yielding the following list of Hecke eigenvalues:

ℓ	3	5	11	13	17	19	23
$a_\ell(f)$	-2	0	0	-4	6	2	0

It can be checked (using Pari, or consulting the tables in [**Cr97**]) that $a_\ell(f)$ is equal to the coefficient $a_\ell(E)$ attached to the (unique, up to isogeny) elliptic curve E of conductor 14 in (6.2), giving a partial numerical verification of theorem 6.5 in this case.

The explicit determination of the harmonic cocycle c encoding the residues of f makes it possible to evaluate numerically the p-adic line integrals attached to f. To illustrate how this can be combined with the theory of complex multiplication to construct algebraic points on E, let $K = \mathbb{Q}(\sqrt{-11})$ be the imaginary quadratic field of smallest discriminant in which both 2 and 7 are inert, and let $\omega_{11} = (1+\sqrt{-11})/2$ denote a generator for its ring of integers. After fixing an embedding $\Psi : \mathcal{O}_K \longrightarrow R$ and letting τ and τ' denote the two fixed points in $\mathcal{H}_7$ of $\iota\Psi(K^\times)$, a computer calculation (carried out to 5 digits of 7-adic accuracy) shows that

$$J = \times\!\!\!\!\!\!\int_{\tau'}^{\tau} f(z)dz = 13149 + 2287\omega_{11} \pmod{7^5}.$$

The image of this integral under the Tate uniformisation attached to E over $\mathbb{Q}_7$ is

$$\Phi_{\mathrm{Tate}}(J) = (10696, 6528 + 9861\omega_{11}) \equiv \left(\frac{7}{11}, \frac{-(41+116\omega_{11})}{121}\right) \pmod{7^5}. \tag{6.5}$$

The latter expression is a global point on $E(K)$.

Although this calculation does not actually prove that equality holds in (6.5), it does illustrate how the ideas presented in the last few chapters can be used in practice to find algebraic points on elliptic curves by p-adic analytic methods. (A second example is detailed in exercise 3.) Calculations of this sort form the basis form the basis for the conjecture of chapter 9 yielding Heegner systems attached to real quadratic fields.

6.6. p-adic L-functions, d'après Schneider-Iovita-Spiess

Recall from chapter 2 that if $f \in S_2(\Gamma_0(N))$ is a normalised eigenform, then one has the corresponding integral representation for its L-series:

$$\Lambda(f,s) := (2\pi)^{-s}\Gamma(s)L(f,s) = \int_0^\infty f(iy)y^{s-1}dy \tag{6.6}$$

which yields the analytic continuation and functional equation for $L(f,s)$. If f is a modular form on a quaternion algebra it is natural to ask for a similar expression describing $L(f,s)$ as a Mellin transform attached to f. On a superficial level, Hecke's construction does not generalise, since it relies on the notion of Fourier expansions which are meaningful only for modular forms attached to the split quaternion algebra $M_2(\mathbb{Q})$.

To arrive at the desired generalisation, it is helpful to view the Mellin transform of f as an integral along the real points of a torus arising from the split quadratic algebra $\mathbb{Q}\times\mathbb{Q} \subset M_2(\mathbb{Q})$. More precisely, let $K \simeq \mathbb{Q}\times\mathbb{Q}$ be the quadratic algebra of diagonal matrices in $M_2(\mathbb{Q})$. This algebra is optimally embedded with respect to the subalgebra $M_0(N)$, in the sense that $K \cap M_0(N) \simeq \mathbb{Z}\times\mathbb{Z}$ is the maximal order of K. The fixed points of $K^\times$ acting by Möbius transformations on $\mathcal{H}^*$ are 0 and ∞. Let z be a rational function with divisor $(0)-(\infty)$. Then $\Lambda(f,s)$ is simply the integral of this function, raised to the power $s-1$, betwen the two fixed points 0 and ∞, against a measure naturally associated to f. By analogy, since a non-split quaternion algebra does not contain the algebra $\mathbb{Q}\times\mathbb{Q}$ as a quadratic subalgebra, one might attempt to define $L(f,s)$ as a Mellin transform of f along the (real, or p-adic) points of a global torus in $B^\times$ coming from the units of a quadratic subalgebra $K \subset B$. Over the reals, this does not seem to lead to a useful analytic object, since the group $(K\otimes\mathbb{R})^\times/\mathcal{O}_K^\times\mathbb{R}^\times$ is a compact group, isomorphic to a circle, which has discrete Pontryagin dual. In the p-adic setting, however, it is natural to consider the space of continuous $\mathbb{C}_p$-valued characters of the compact p-adic group $(K\otimes\mathbb{Q}_p)^\times/\mathcal{O}_K^\times\mathbb{Q}_p^\times$. This "$p$-adic dual" is not discrete and is in fact endowed with a non-trivial topological and p-adic analytic structure.

To simplify our discussion, we will only treat the case where the prime p is inert in $K/\mathbb{Q}$. (The case where p is split in K is treated in exercise 4.) In that case the group $\iota(K_p^\times)$ acting on $\mathcal{H}_p$ has two fixed points α and $\bar\alpha$ which belong to $\mathcal{H}_p \cap K$ and are interchanged under the action of $\mathrm{Gal}(K_p/\mathbb{Q}_p)$. Let $\left(\frac{z-\alpha}{z-\bar\alpha}\right)$ be a rational function with divisor $(\alpha)-(\bar\alpha)$. It is tempting to define, by analogy with (6.6)

$$L_p(f,s) \text{``}\overset{?}{=}\text{''} \int_{\bar\alpha}^{\alpha} f(z)\left(\frac{z-\alpha}{z-\bar\alpha}\right)^{s-1} dz. \tag{6.7}$$

If s belongs to $\mathbb{Z}$, then $\omega_g = f(z)\left(\frac{z-\alpha}{z-\bar\alpha}\right)^{s-1} dz$ is a rigid analytic differential form on $\mathcal{H}_p - \{\alpha,\bar\alpha\}$ with associated boundary distribution on $\mathbb{P}_1(\mathbb{Q}_p)$ given by

$$\mu_g(t) = \mu_f(t)\left(\frac{t-\alpha}{t-\bar\alpha}\right)^{s-1}.$$

Although the integral appearing in (6.7) does not converge, if we simply *ignore* the terms arising from the end points α and $\bar\alpha$ we obtain the following candidate for a generalisation of (6.6), in which the argument of K has been inserted to emphasize the essential dependence of this definition on the chosen quadratic subalgebra:

$$L_p^{\mathrm{naive}}(f,K,s) := \int_{\mathbb{P}_1(\mathbb{Q}_p)} \log\left(\frac{t-\alpha}{t-\bar{\alpha}}\right)\left(\frac{t-\alpha}{t-\bar{\alpha}}\right)^{s-1} d\mu_f(t). \tag{6.8}$$

This expression does converge, and even interpolates to $s \in \mathbb{Z}_p$. Moreover, it is the derivative with respect to s of an even simpler expression:

DEFINITION 6.14. The Schneider-Iovita-Spiess L-function attached to f and K is the expression

$$L_p(f,K,s) := \int_{\mathbb{P}_1(\mathbb{Q}_p)} \left(\frac{t-\alpha}{t-\bar{\alpha}}\right)^{s-1} d\mu_f(t).$$

REMARK 6.15. This type of definition was proposed by Schneider in [**Sch84**], with the role of K being played by the local split algebra $\mathbb{Q}_p \times \mathbb{Q}_p$ embedded in $B_p \simeq M_2(\mathbb{Q}_p)$. Since this embedding has no global origin it is unclear what relation (if any) Schneider's construction bears with classical special values and with other types of p-adic L-functions. The idea of requiring that K arise from a global quadratic subalgebra of B (thus giving more rigidity to Schneider's construction) was arrived at independently by Iovita and Spiess. (Cf. for example [**BDIS02**] for a more thorough discussion.)

The p-adic Mellin transform $L_p(f,K,s)$ is a more tractable object than its classical counterpart of (6.6), as the following theorem illlustrates. Let E^K be the elliptic curve over $\mathbb{Q}$ obtained by twisting E by the quadratic character attached to K.

THEOREM 6.16. *The order of vanishing of $L_p(f,K,s)$ is greater than both the rank of $E(\mathbb{Q})$ and the rank of $E^K(\mathbb{Q})$.*

This theorem lies beyond the scope of our discussion. A proof, as well as a conjecture describing the precise order of vanishing of $L_p(f,K,s)$ and the philosophy underlying it, is given in [**BD03**].

REMARK 6.17. In contrast, the mechanism whereby large rank of $E(\mathbb{Q})$ forces extra vanishing in the complex L-function $L(f,s)$ is not understood at all. To take stock of the ignorance surrounding this question, note that the following remains open: Does there exist a curve E for which $\mathrm{ord}_{s=1}L(E,s) > 3$? On the other hand theorem 6.16 yields examples of elliptic curves for which one can show that $\mathrm{ord}_{s=1}L_p(E,K,s) > 23$, thanks to the elliptic curves curves of large rank that have been produced by Mestre and his school.

6.7. A Gross-Zagier formula

Suppose for simplicity that K has class number one and that

$$\Psi(K) \cap R = \Psi(\mathcal{O}), \qquad \text{where } \mathcal{O} = \mathcal{O}_K[1/p].$$

Let f be a modular eigenform on $\mathcal{H}_p/\Gamma^{(p)}_{N^+,N^-}$ with integer residues and let E be the strong Weil curve associated to it. The following gives an arithmetic interpretation of the special value $L_p'(f,K,1)$ in the spirit of the Gross-Zagier formula (theorem 3.20 of chapter 3).

THEOREM 6.18. *There exists a global point $P_K \in E(K)$ such that*

$$\beta\Phi_{\mathrm{Tate}}(\exp(L_p'(f,K,1))) = P_K \pmod{E(K_p)_{\mathrm{tors}}}.$$

PROOF. Let α and $\bar{\alpha}$ be the fixed points of $\iota\Psi(K^\times)$ acting on $\mathcal{H}_p$, and let

$$P_K = \Phi^{(p)}_{N^+,N^-}((\alpha) - (\bar{\alpha})) \in E(\mathbb{C}_p).$$

Since $(\alpha) - (\bar{\alpha})$ belongs to $\mathrm{Div}^0(CM(\mathcal{O}))$, and K is its own Hilbert class field, it follows from theorem 6.11 that P_K belongs to $E(K)$. On the other hand,

$$L'_p(f, K, 1) = \int_{\mathbb{P}_1(\mathbb{Q}_p)} \log\left(\frac{t-\alpha}{t-\bar{\alpha}}\right) d\mu_f(t) = \int_{\bar{\alpha}}^{\alpha} f(z)dz.$$

Hence

$$\exp(L'_p(f, K, 1)) = \times\!\!\!\!\!\!\int_{\bar{\alpha}}^{\alpha} f(z)dz \pmod{(K_p^\times)_{\mathrm{tors}}}.$$

Applying the Tate uniformisation and the isogeny β to both sides yields the desired result, in light of (6.1). □

Further results

The first proof of the Čerednik-Drinfeld theorem was obtained by Čerednik [**Ce76**] building on work of Ihara [**Ih68**]. Drinfeld's proof [**Dr76**] which also gives a moduli interpretation to the p-adic upper half plane is explained in [**BC92**].

The numerical example studied in section 6.5 is taken from [**BD96**]. More details and the general context for the Schneider-Iovita-Spiess approach to p-adic L-functions are explained in [**BD01**] and [**BDIS02**]. The proof of theorem 6.16 (which follows from a more general Iwasawa-theoretic "main conjecture" for certain p-adic L-functions which include the rigid analytic L-function discussed above as a special case) is given in [**BD03**].

Exercises

(1) Let A be an abelian surface with quaternionic multiplication by R, so that R acts on A by endomorphisms which are defined over $\mathbb{Q}$. We say that A has *complex multiplication* by an order $\mathcal{O}$ in a quadratic field K if there is an inclusion $\mathcal{O} \subset \mathrm{End}_R(A)$, where $\mathrm{End}_R(A)$ denoted the algebra of endomorphisms of A which commute with R. Show that if A has complex multiplication by $\mathcal{O}$, it is isomorphic over $\mathbb{Q}$ to a product $E \times E$ of elliptic curves with complex multiplication by $\mathcal{O}$.

(2) Let B be the algebra of Hamilton's quaternions over $\mathbb{Q}$, and let

$$R_0 = \mathbb{Z}[i, j, k, \frac{1+i+j+k}{2}]$$

be Hurwitz's maximal order. For any odd prime p, let $R = R_0[1/p]$, choose an identification ι of $B \otimes \mathbb{Q}_p$ with $M_2(\mathbb{Q}_p)$, and set $\Gamma = \iota(R_1^\times)$. Show that $\Gamma\backslash\mathcal{T}_p$ is a graph with two vertices. Give an example of a group $\Gamma \subset B^\times$ for which the quotient $\Gamma\backslash\mathcal{T}_p$ has more than two vertices.

(3) There are two isogeny classes of elliptic curves of conductor 26, which are labelled $26A$ and $26B$ in the Cremona tables; their minimal Weierstrass equations are

$$26A : y^2 + xy + y = x^3 - 5x - 8, \quad 26B : y^2 + xy + y = x^3 - x^2 - 3x + 3.$$

Let B and R_0 be as in section 6.5, and let $R = R_0[1/13]$. Choose an embedding ι of B_{13} into $M_2(\mathbb{Q}_{13})$ by setting

$$\iota(i) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \iota(j) = \begin{pmatrix} \rho & \rho+1 \\ \rho+1 & -\rho \end{pmatrix}, \quad \iota(k) = \begin{pmatrix} \rho+1 & -\rho \\ -\rho & -\rho-1 \end{pmatrix},$$

where $\rho = \lim_{n\longrightarrow\infty} 3^{13^n}$ is a primitive cube root of unity in $\mathbb{Z}_{13}$. Let $\Gamma = \iota(R_1^\times)$ and let $\mathcal{T}_{13} = \mathcal{T}_0 \cup \mathcal{T}_1$ be the Bruhat-Tits tree of $\mathbf{PGL}_2(\mathbb{Q}_{13})$.

(a) Show that there are exactly two orbits for the action of Γ on $\mathcal{T}_0$, and that there are three orbits for the action of Γ on $\mathcal{T}_1$. More precisely show that these three orbits E_1, E_2 and E_3 are characterised by the property

$$e \in \begin{cases} E_1 & \text{if } \ e \text{ is } \Gamma\text{-equivalent to } e_\infty, e_0, e_3, \text{ or } e_4, \\ E_2 & \text{if } \ e \text{ is } \Gamma\text{-equivalent to } e_1, e_7, e_{11}, \text{ or } e_{12}, \\ E_3 & \text{if } e \text{ is } \Gamma\text{-equivalent to } e_2, e_5, e_6, e_8, e_9, \text{ or } e_{10}. \end{cases}$$

Draw the quotient graph $\mathcal{T}_{13}/\Gamma$.

(b) Show that the space of Γ-invariant harmonic cocycles on the tree $\mathcal{T}_{13}$ is two-dimensional, by showing that such a harmonic cocycle is completely determined by its values on the ordered edges e'_∞, e'_1 and e'_2 associated to e_∞, e_1 and e_2 and having v_o as source.

(c) Compute the action of the Hecke operator T_3 on a basis of the space of Γ-invariant harmonic cocycles on $\mathcal{T}_{13}$. Show that a basis of eigenvectors for this action is given by the harmonic cocycles c_1 and c_2, where

$$c_1(e'_\infty) = 1, \qquad c_1(e'_1) = -1, \qquad c_1(e'_2) = 0.$$

$$c_2(e'_\infty) = 3, \qquad c_2(e'_1) = 3, \qquad c_2(e'_2) = -4.$$

(d) Let f_j be the rigid analytic modular form in $S_2(\Gamma)$ attached to the cocycle c_j. Show that f_1 is associated to the elliptic curve $26A$ and that f_2 to the elliptic curve $26B$ in the correspondence of theorem 6.5.

(e) Use the theory of complex multiplication to find points on these elliptic curves defined over $\mathbb{Q}(\sqrt{-11})$ by evaluating the appropriate 13-adic integral.

(4) This exercise studies the Schneider-Iovita-Spiess p-adic L-function $L_p(f, K, s)$ in the case where p is split in the quadratic imaginary field $K \subset B$.

(a) Show that the two fixed points α and $\bar{\alpha}$ of $\iota(K^\times)$ acting on $\mathbb{P}_1(\mathbb{C}_p)$ by Möbius transformations belong to the boundary $\mathbb{P}_1(\mathbb{Q}_p)$ of $\mathcal{H}_p$.

(b) If $\mathcal{O}$ is a $\mathbb{Z}[1/p]$-order of K, show that $\mathcal{O}_1^\times$ is of rank one. Let ϵ be a generator for this group.

(c) Show that the subgroup of $\Gamma = \iota(R_1^\times)$ which fixes α and $\bar{\alpha}$ is of rank one. Let $\gamma \in \mathbf{SL}_2(\mathbb{Q}_p)$ be a generator of this group.

(d) Fix an embedding of K into $\mathbb{Q}_p$ and let ϵ_p be the image of ϵ in $\mathbb{Q}_p$. Let $\log : \mathbb{Q}_p^\times \longrightarrow \mathbb{Q}_p$ be a branch of the p-adic logarithm, chosen so that $\log(\epsilon_p) = 0$. Define $x^s = \exp(s\log(x))$. Show that the measure

$$\mu(t) = \left(\frac{t-\alpha}{t-\bar{\alpha}}\right)^{s-1} d\mu_f(t)$$

is well-defined on $\mathbb{P}_1(\mathbb{Q}_p) - \{\alpha, \bar{\alpha}\}$ and is invariant under the action of γ.

(e) Let D be a fundamental region for the action of γ on $\mathbb{P}_1(\mathbb{Q}_p) - \{\alpha, \bar{\alpha}\}$. Show that the expression

$$L_p(f, K, s) := \int_D \left(\frac{t-\alpha}{t-\bar{\alpha}}\right)^{s-1} d\mu_f(t)$$

converges for $s \in \mathbb{Z}_p$ and is independent of the choice of D.

(f) Show that

$$L_p'(f, K, 1) = \int_\tau^{\gamma\tau} f(z)dz,$$

where τ is any element of $\mathcal{H}_p$. Show that this expression is a rational multiple of $\log(q)$, where q is the Tate period attached to $E_{/\mathbb{C}_p}$.

(5) Prove lemma 6.10.

CHAPTER 7

Totally real fields

To a modular elliptic curve E over $\mathbb{Q}$ of conductor N three different types of modular parametrisation have been attached so far:

- The classical modular curve parametrisation
$$\Phi_N : \mathcal{H}/\Gamma_0(N) \longrightarrow E(\mathbb{C})$$
of chapter 2;
- The Shimura curve parametrisation
$$\Phi_{N^+,N^-} : \mathrm{Div}^0(\mathcal{H}/\Gamma_{N^+,N^-}) \longrightarrow E(\mathbb{C})$$
of chapter 5;
- The rigid analytic parametrisation
$$\Phi^{(p)}_{N^+,N^-} : \mathrm{Div}^0(\mathcal{H}_p/\Gamma^{(p)}_{N^+,N^-}) \longrightarrow E(\mathbb{C}_p)$$
arising from the theory of Čerednik and Drinfeld.

In all cases it has been possible to develop a notion of CM points on the appropriate (complex or p-adic) upper half-plane, whose image under the corresponding modular parametrisation yields points on E defined over abelian extensions of certain quadratic *imaginary* fields.

In this chapter we turn to the question of what happens if the ground field $\mathbb{Q}$ is replaced by a more general number field F. This question is not motivated merely by the pursuit of generalisation. Rather, an examination of the number field case suggests a broader perspective on modular parametrisations, a perspective that will be germane to the conjectural theory of Heegner points attached to real quadratic fields presented in chapter 9.

7.1. Elliptic curves over number fields

Let E be an elliptic curve over over a number field F. The *conductor* of E is now an integral ideal $\mathcal{N}$ of $\mathcal{O}_F$. Let $|\mathfrak{n}|$ denote the norm of the (fractional or integral) ideal $\mathfrak{n}$. For each prime $\mathfrak{p}$ of F, define an integer $a(\mathfrak{p})$ by the rule:

$$a(\mathfrak{p}) = 1 + |\mathfrak{p}| - \#E(\mathcal{O}_F/\mathfrak{p}), \quad \text{if } \mathfrak{p} \not| N, \tag{7.1}$$

and $a(\mathfrak{p}) = 0$ (resp. 1, -1) if E has additive (resp split, non-split multiplicative) reduction at $\mathfrak{p}$. To f and N is associated a space of "automorphic forms"

$$S_2(N) \subset L^2(\mathbf{GL}_2(F)\backslash\mathbf{GL}_2(\mathbb{A}_F)),$$

equipped as in the case where $F = \mathbb{Q}$ with an action of Hecke operators indexed by the primes of F.

We will not go here into the details of the precise definition of this space of automorphic forms, as this would take us too far afield.

A generalisation of the Shimura-Taniyama-Weil conjecture predicts that there exists an automorphic form $f \in S_2(N)$ such that

$$T_\ell f = a_\ell(E) f \text{ for all primes } \ell \text{ of } F \text{ not dividing } N.$$

Our hope (formulated vaguely here, and made more precise in the next three chapters) is that the existence of such an f provides a handle on the arithmetic of $E_{/F}$, useful not just in proving analyticity properties for the L-function $L(E/F, s)$ and algebraicity results for its special values, but also in constructing algebraic points on E, as in the prototypical case where $F = \mathbb{Q}$. Essential difficulties arise because there is no systematic generalisation of the notion of modular curve in the number field setting. For instance, here is what happens in the simplest (and perhaps most interesting) case where F is a quadratic field:

1. In the case (discussed more thoroughly in the next sections) where F is real, the form f is a *Hilbert modular form* defined on a Hilbert modular surface. An analogue of the Eichler-Shimura construction can often be given by applying the Jacquet-Langlands correspondence and passing to a form on an appropriate Shimura curve, but this is not always possible: for instance it fails when f is attached to an elliptic curve with everywhere good reduction over F. The theory presented in this and the next chapter proposes a construction of algebraic points on E directly in terms of the periods of f which would apply to all cases where F is real quadratic.

2. If F is imaginary quadratic, the form f corresponds to a differential form on the upper-half space $\mathbb{C} \times \mathbb{R}^{>0}$ invariant under the action of a discrete arithmetic subgroup $\Gamma \subset \mathbf{SL}_2(\mathbb{C})$ (a so-called *Bianchi group*, cf. [**EGM98**]). By obvious dimension considerations, the hyperbolic three-manifold on which f is defined cannot correspond to the complex points of an algebraic variety and one is at a loss to propose, even conjecturally, an appropriate generalisation of the Eichler-Shimura construction in this setting (in spite of [**HST93**], [**T94**] which succeeds in attaching to f a compatible system of ℓ-adic Galois representations having the same properties as those attached to E). The case where F is imaginary quadratic is not touched upon at all in these notes, although a natural extension of the theory proposed in chapter 9 coud be expected to yield conjectural p-adic analytic constructions of non-trivial Heegner systems attached to elliptic curves defined over such fields.

7.2. Hilbert modular forms

We specialise the discussion – and are correspondingly more precise – in the case where F is a totally real field of degree $n+1$ over $\mathbb{Q}$. Let $\iota_0, \ldots, \iota_n$ be the distinct real embeddings of F. The $(n+1)$-tuple of embeddings $\iota = (\iota_0, \ldots, \iota_n)$ induces embeddings

$$F \longrightarrow \mathbb{R}^{n+1}, \quad M_2(F) \longrightarrow M_2(\mathbb{R})^{n+1}, \quad \mathbf{PSL}_2(F) \longrightarrow \mathbf{PSL}_2(\mathbb{R})^{n+1},$$

which will all be denoted by ι by a slight abuse of notation. If γ is an element of F, or a matrix with entries in F, it will occasionally be convenient to write γ_j or $\iota_j(\gamma)$ for the image of γ under the j-th real embedding of F.

Note that the group $\mathbf{SL}_2(\mathbb{R})^{n+1}$ acts naturally on the product $\mathcal{H}_0 \times \mathcal{H}_1 \times \cdots \times \mathcal{H}_n$ of $(n+1)$ copies of the complex upper half-plane indexed in the same way as the places of F. Given $\underline{\tau} = (\tau_0, \ldots, \tau_n) \in \mathcal{H}^{n+1}$ and $\underline{\gamma} = (\gamma_0, \ldots, \gamma_n) \in \mathbf{SL}_2(\mathbb{R})^{n+1}$, we will write $\underline{\gamma}\underline{\tau} = (\gamma_0\tau_0, \ldots, \gamma_n\tau_n)$.

LEMMA 7.1. *The group* $\Gamma = \iota(\mathbf{PSL}_2(\mathcal{O}_F))$ *acts discretely on* $\mathcal{H}^{n+1}$.

PROOF. Note that the image of $\mathcal{O}_F$ under ι is discrete in $\mathbb{R}^{n+1}$. Hence so is the image of $M_2(\mathcal{O}_F)$ in $M_2(\mathbb{R}^{n+1})$, and therefore the group $\Gamma := \iota(\mathbf{PSL}_2(\mathcal{O}_F))$ is a discrete subgroup of $\mathbf{PSL}_2(\mathbb{R})^{n+1}$. Since $\mathcal{H}^{n+1}$ is the quotient of $\mathbf{PSL}_2(\mathbb{R})^{n+1}$ by the compact subgroup $\mathbf{O}_2(\mathbb{R})^{n+1}$ the result follows by the same argument that was used in the proof of lemma 4.6 of chapter 4. □

In order to work as much as possible using classical rather than adèlic notations, it is useful to make the following simplifying assumption:

HYPOTHESIS 7.2. *The field F has narrow class number one.*

Thus it is assumed that every ideal of $\mathcal{O}_F$ is principal and has a totally positive generator, so that there exist units $\epsilon_0, \epsilon_1, \ldots, \epsilon_n$ with the property that

$$\iota_k(\epsilon_j) > 0, \text{ if } k \neq j, \quad \iota_j(\epsilon_j) < 0.$$

DEFINITION 7.3. A *Hilbert modular form* of weight $(k_0, \ldots, k_n)$ on Γ is a holomorphic function

$$f : \mathcal{H}_0 \times \cdots \times \mathcal{H}_n \longrightarrow \mathbb{C}$$

such that

$$f(\underline{\gamma}\underline{\tau}) = (c_0\tau_0 + d_0)^{k_0} \cdots (c_n\tau_n + d_n)^{k_n} f(\underline{\tau}) \text{ for all } \underline{\gamma} = \left(\begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix} \right)_j \in \Gamma.$$

Of special relevance to elliptic curves is the case where

$$\Gamma = \iota(\Gamma_0(N)), \text{ where } \Gamma_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathcal{O}_F) \text{ such that } N|c \right\},$$

and $(k_0, \ldots, k_n) = (2, \ldots, 2)$. The Hilbert modular form f is then said to be of *parallel weight* 2 with respect to Γ, and the space of all such forms is denoted $S_2(N)$, as in the case where $F = \mathbb{Q}$.

The theory of Hilbert modular forms exhibits many features analogous to the classical situation where $F = \mathbb{Q}$. Most notable perhaps is the existence of Fourier expansions. More precisely, since every matrix of the form $\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ with $\lambda \in \mathcal{O}_F$ belongs to $\Gamma_0(N)$, we have:

$$f(\underline{\tau} + \underline{\lambda}) = f(\underline{\tau}), \text{ for all } \lambda \in \mathcal{O}_F.$$

Therefore f is periodic under translation by elements of $\iota(\mathcal{O}_F)$ and admits a Fourier expansion

$$f(\underline{\tau}) = \sum_{\nu \in \mathcal{O}_F} a_\nu(f) e^{2\pi i (\underline{\tau} \cdot (\underline{\nu}/\underline{d}))},$$

where d is a totally positive generator of the different of F, and

$$\underline{\tau} \cdot (\underline{\nu}/\underline{d}) := \tau_0\nu_0/d_0 + \tau_1\nu_1/d_1 + \cdots + \tau_n\nu_n/d_n.$$

The Fourier coefficients $a_\nu(f)$, indexed by elements ν of $\mathcal{O}_F$, satisfy the following basic properties:

- The Koecher principle implies that, if $\iota_j(\nu) < 0$ for some j, then $a_\nu(f) = 0$: in other words, $a_\nu(f)$ is non-zero only on totally positive elements. (Cf. for example [**Gar90**] §1.4.)

- The invariance of f under the matrices of the form $\begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix}$, where ϵ is a unit in $\mathcal{O}_F$, implies that $a_\nu(f) = a_{\epsilon^2\nu}(f)$ for all such units. Under the condition that F has narrow class number one, every totally positive unit of $\mathcal{O}_F$ is a square of a unit and hence $a_\nu(f)$ (for ν totally positive) depends only on the ideal generated by ν, not on ν itself: thus $\nu \mapsto a_\nu(f)$ can be viewed as a function on the ideals of $\mathcal{O}_F$.

As in the definitions of chapter 2, it is possible to define the Fourier expansion of f at an arbitrary cusp $\gamma^{-1}\infty$ (with $\gamma \in \mathbf{SL}_2(\mathcal{O}_F)$) by expanding $f|_\gamma$ as a Fourier series. Let a_0^γ denote the constant coefficient in this expansion.

DEFINITION 7.4. If $a_0^\gamma(f) = 0$, for all $\gamma \in \mathbf{SL}_2(\mathcal{O}_F)$, then f is said to be a *cusp form.*

We shall denote by $S_2(N)$ the space of cusp forms of parallel weight $(2, \ldots, 2)$ on $(\mathcal{H}_0 \times \cdots \times \mathcal{H}_n)/\Gamma_0(N)$. The space $S_2(N)$ is equipped with many of the familiar structures present in the case where $F = \mathbb{Q}$, such as the existence of an inner product, and a collection of mutually commuting self-adjoint Hecke operators T_ℓ indexed by the primes $\ell \not| N$ of $\mathcal{O}_F$, leading to a definition of a *simultaneous eigenform* for these Hecke operators. If f is such an eigenform, then $a_\ell(f)$ is equal to the eigenvalue of the Hecke operator T_ℓ acting on f, as in the classical case. For more details the reader is invited to consult the book of Bump [**Bu97**].

7.3. The Shimura-Taniyama-Weil conjecture

Recall the running assumption that F is a totally real field of narrow class number one. If E is an elliptic curve over F, the definition of the coefficients $a(\mathfrak{p})$ attached to E given in equation (7.1) can be extended to all integral ideals in the usual way through the equality of Dirichlet series

$$\prod_{\lambda \not| N}(1 - a(\lambda)|\lambda|^{-s} + |\lambda|^{1-2s})^{-1} \prod_{\lambda | N}(1 - a(\lambda)|\lambda|^{-s})^{-1} = \sum_{\nu} a(\nu)|\nu|^{-s},$$

where the products are taken over prime ideals and the sum is taken over integral ideals of F. We may also view $a(\nu)$ as a function on the totally positive elements of $\mathcal{O}_F$ in the obvious way.

The natural extension of the Shimura-Taniyama conjecture for elliptic curves over totally real fields can be formulated concretely as follows:

CONJECTURE 7.5. *Let E be an elliptic curve of conductor N over F. Then the series defined by*

$$f(\underline{\tau}) = \sum_{\nu \gg 0} a(\nu) e^{2\pi i(\underline{\tau} \cdot (\underline{\nu}/\underline{d}))}$$

is an eigenform in $S_2(N)$.

The ideas of Wiles, as generalised by a number of other mathematicians, notably Diamond and Skinner, lead to the proof of many special cases of the Shimura-Taniyama-Weil conjecture over totally real fields, as the following result indicates:

THEOREM 7.6 (Skinner-Wiles). *Let E be a semistable elliptic curve of conductor N over a totally real field F. Assume further that*

- *Either $[F : \mathbb{Q}]$ is odd, or $N \neq 1$.*
- *The extension of F generated by the coordinates of the 3-division points of E has Galois group $\mathbf{GL}_2(\mathbb{F}_3)$ over F.*

Then conjecture 7.5 holds for E.

The semi-stability assumption and the second condition are present for technical reasons having to do with difficulties that are encountered in certain lifting theorems of Wiles when the mod 3 Galois representation attached to E is reducible. The first condition arises from a more serious (and, from the point of view of this monograph, more interesting) difficulty that is not present in the case of curves over $\mathbb{Q}$: the Eichler-Shimura construction does not always admit a suitable generalisation to the context of totally real fields. This question is discussed further in the next section.

7.4. The Eichler-Shimura construction for totally real fields

THEOREM 7.7. *Let f be a Hilbert modular eigenform in $S_2(N)$ with rational Hecke eigenvalues. Assume that $[F:\mathbb{Q}]$ is odd, or that there is a prime $\mathfrak{p}$ of F dividing N exactly. Then there is an elliptic curve E/F of conductor N such that*

$$a_\lambda(E) = a_\lambda(f), \text{ for all } \lambda \nmid N.$$

SKETCH OF THE CONSTRUCTION. Let B be the quaternion algebra over F ramified precisely at the archimedean places $v_1, \ldots, v_n$, if $n+1 = [F:\mathbb{Q}]$ is odd, or at $v_1, \ldots, v_n, \mathfrak{p}$, if $n+1$ is even. Let R be an Eichler order of level N (resp. $N/\mathfrak{p}$) in B if $[F:\mathbb{Q}]$ is odd (resp. even). Since B is split at the place v_0, one may choose an identification

$$\iota_0 : B \otimes_{v_0} \mathbb{R} \longrightarrow M_2(\mathbb{R}).$$

Let $\Gamma := \iota_0(R_1^\times)$. This group acts discretely on $\mathcal{H}$ with compact quotient and $\mathcal{H}/\Gamma$ can be interpreted as the complex points of a *Shimura curve* X which is defined over F. An appropriate generalisation of the theory of Jacquet-Langlands to the context of modular forms on B produces a modular form $g \in S_2(\mathcal{H}/\Gamma)$ attached to f. The elliptic curve E_f is constructed analytically as a quotient of $\mathrm{Jac}(X)$ following a generalisation to totally real fields of the theory alluded to in section 4.4. For a detailed description of the theory of Shimura curves attached to quaternion algebras over totally real fields, and their use in defining a variant of the Eichler-Shimura construction in this context, see chapters 1 and 3 of [**Zh01a**] and the references contained therein. $\square$

The cases where the Eichler-Shimura construction can be carried out yield a Shimura curve parametrisation

$$\mathrm{Div}^0(\mathcal{H}/\Gamma) \longrightarrow E(\mathbb{C}),$$

where Γ is the group introduced in the discussion of theorem 7.7. It should be noted, however, that not all elliptic curves over a totally real field F can be uniformised by a Shimura curve in this way. For instance, the theory excludes elliptic curves with everywhere good reduction defined over a totally real field of even degree.

7.5. The Heegner construction

The Heegner point construction generalises precisely to the situations where E is the quotient of the Jacobian of a Shimura curve, i.e., can be obtained from the Eichler-Shimura construction.

For in that case, letting Γ be the subgroup of $\mathbf{SL}_2(\mathbb{R})$ arising as in the previous section from a quaternion algebra B split at precisely one archimedean place v_0 of F, define for each $\tau \in \mathcal{H}/\Gamma$ the associated order of τ to be

$$\mathcal{O}_\tau := \{\gamma \in R \text{ such that } \det(\gamma) \neq 0 \text{ and } \gamma\tau = \tau\} \cup \{0\}.$$

LEMMA 7.8. *The order $\mathcal{O}_\tau$ is either equal to $\mathcal{O}_F$, or to an order in a quadratic CM extension of F.*

PROOF. By the same reasoning as in section 3.2, the order $\mathcal{O}_\tau$ is a commutative subring of B containing $\mathcal{O}_F$. Its fraction field K is therefore a commutative subfield of B, hence is either equal to F or to a quadratic extension of F. Suppose the latter occurs. For all the archimedean places v_j of F, one has $K \otimes_{v_j} \mathbb{R} = \mathbb{R} \times \mathbb{R}$ or $\mathbb{C}$. The former cannot happen for $j = 1, \ldots, n$ since $B \otimes_{v_j} \mathbb{R} = \mathbb{H}$ has no zero divisors. It is also the case that $K \otimes_{v_0} \mathbb{R} = \mathbb{C}$ because the action of $K^\times$ on $\mathcal{H}_0$ has a fixed point. □

Given any order $\mathcal{O}$ in a CM extension K of F, denote by $CM(\mathcal{O})$ the set of $\tau \in \mathcal{H}/\Gamma$ whose associated order $\mathcal{O}_\tau$ is equal to $\mathcal{O}$, and denote by H the ring class field of K attached to $\mathcal{O}$.

THEOREM 7.9. *If τ belongs to $CM(\mathcal{O})$, then $\Phi_{N^+,N^-}(\mathrm{Div}^0(CM(\mathcal{O})))$ is contained in $E(H)$.*

REMARKS ON THE PROOF. The idea here is to give a moduli interpretation to the Shimura curve $\mathcal{H}/\Gamma$ (or at least to a finite covering of it) and thereby interpret the CM points on this quotient as coming from the moduli of certain abelian varieties with CM by $\mathcal{O}$ which are then defined over the appropriate ring class field. See ch. 2 of [**Zh01a**] for more details. □

The occasional absence of a Shimura curve parametrisation raises the problem of constructing algebraic points on E in such situations - for example, when E is an elliptic curve over a totally real field F of even degree. Note that for such an E we have:

PROPOSITION 7.10. *Assume conjecture 3.19. If K is any CM extension of F, then* $\mathrm{sign}(E, K) = 1$.

PROOF. For under these hypotheses the analytic set $S_{E,K} = \{v_0, \ldots, v_n\}$ consists only of the archimedean places of F (viewed as complex places of K). Since $n + 1$ is even, the result follows from conjecture 3.19. □

It follows that if E is an elliptic curve with everywhere good reduction over a totally real field of even degree, then one should expect no Heegner systems attached to any CM extension of F, so that the theory of complex multiplication is presumably unable to produce any Heegner system attached to E – a difficulty which runs parallel to the fact that E does not appear in the Jacobian of any Shimura curve over F.

7.6. A preview of chapter 8

Let us return to the setting where f is a Hilbert modular form of parallel weight 2 on $(\mathcal{H}_0 \times \cdots \times \mathcal{H}_n)/\Gamma_0(N)$ attached to the elliptic curve E over F. Given any $\tau \in \mathcal{H}_0$, write Γ_τ for the stabiliser of τ in $\Gamma = \Gamma_0(N)$.

LEMMA 7.11. *The group Γ_τ is an abelian group of rank at most n.*

PROOF. Let $\mathcal{O}_\tau \subset M_0(N)$ be the associated order of τ in $M_0(N)$, i.e., the algebra of matrices which preserve the line spanned by the column vector $(\tau, 1)$. Then $\mathcal{O}_\tau$ is isomorphic either to $\mathcal{O}_F$ or to an $\mathcal{O}_F$-order in a quadratic extension K of F (i.e,. a subring of K which contains $\mathcal{O}_F$, is finitely generated as an $\mathcal{O}_F$-module, and generates K as a $\mathbb{Q}$-algebra). In the former case Γ_τ is trivial, and in the latter case it is identified with the group of elements of ${\mathcal{O}_\tau}^\times$ whose norm from K to F is equal to one. Since the torus $K^\times \subset \mathbf{GL}_2(F)$ has a fixed point on $\mathcal{H}_0$, it follows that $K \otimes_{v_0} \mathbb{R} \simeq \mathbb{C}$, so that the place v_0 lies below a complex place of K. Hence K has at most $2n+1$ archimedean places; by the Dirichlet unit theorem

$$\mathrm{rank}_{\mathbb{Z}}(\mathcal{O}_\tau^\times) \leq 2n; \quad \mathrm{rank}_{\mathbb{Z}}(\mathcal{O}_F^\times) = n.$$

Since the norm map $\mathcal{O}_\tau^\times \longrightarrow \mathcal{O}_F^\times$ has finite cokernel, the result follows. $\square$

REMARK 7.12. Equality is attained in lemma 7.11 precisely when K is an extension of F which is complex at v_0 and real at all the other archimedean places of F. Such an extension will be called an *almost totally real* (ATR) extension of F. This notion depends on the chosen ordering $v_0, \ldots, v_n$ of the real embeddings of F, or at least on singling out the distinguished place v_0. If K is an ATR extension, we will customarily fix as part of the data an extension of v_0 to a complex embedding of K, making it possible to view K as a subfield of the complex numbers.

DEFINITION 7.13. An *ATR point* is a point $\tau \in \mathcal{H}_0$ such that $\mathrm{rank}(\Gamma_\tau) = n$. Equivalently, τ is ATR if it belongs to $\mathcal{H}_0 \cap K$ for some ATR extension K of F.

Let $\mathcal{H}_0'$ denote the set of ATR points on $\mathcal{H}_0$.

The basic insight to be developed in the next chapter is that there ought to be a *natural substitute*

$$\Phi_N' : \mathcal{H}_0'/\Gamma \longrightarrow E(\mathbb{C})$$

for the Weil uniformisation of chapter 2. This new type of uniformisation, as in the case $n = 0$ which it generalises, is constructed from periods of the differential $(n+1)$-form

$$\omega_f := (2\pi i)^{n+1} f(\underline{\tau}) d\underline{\tau}$$

attached to E. The precise definition of Φ_N', based on cohomological properties of the groups Γ and Γ_τ, is given in the next chapter. The main conjecture that will emerge, a natural generalisation of the theory of Heegner points, states roughly that if $\tau \in \mathcal{H}_0 \cap K$ is an ATR point, then $\Phi_N'(\tau)$ is a global point defined over an appropriate ring class field of K.

Further results

The book of Bump [**Bu97**], particularly the first chapter, provides an excellent introduction to the theory of Hilbert modular forms. Other good references are the books by Garrett [**Gar90**] and Freitag [**Frei90**].

A useful account of the theory of Shimura curves attached to totally real fields, and of Heegner points arising from such Shimura curve parametrisations is the article of Zhang [**Zh01a**] which also proves the generalisation of the formula of Gross and Zagier in this context.

CHAPTER 8

ATR points

We retain the setting of the previous chapter. Thus F is a totally real field of narrow class number 1 and of degree $n+1$ over $\mathbb{Q}$, and E is an elliptic curve over F of conductor N, attached to a Hilbert modular form $f \in S_2(N)$ of level N and parallel weight two. Let

$$\omega_f = (2\pi i)^{n+1} f(\tau_0, \dots, \tau_n) d\tau_0 \cdots d\tau_n$$

be the holomorphic dfferential $(n+1)$-form on $(\mathcal{H}_0 \times \cdots \times \mathcal{H}_n)/\Gamma$ attached to f.

8.1. Period integrals

For $j = 0, \dots, n$, let x_j and $y_j \in \mathcal{H}_j$ be points on the j-th upper half-plane indexed by the real embeddings ι_j of F. Write

$$\int_{x_0}^{y_0} \int_{x_1}^{y_1} \cdots \int_{x_n}^{y_n} \omega_f \in \mathbb{C} \tag{8.1}$$

for the usual multiple integral attached to the diferential form ω_f. It can be calculated numerically by exploiting the Fourier expansion of f, and is given by the expression

$$|d| \sum_{\nu \gg 0} a(\nu)/|\nu| \prod_{j=0}^{n} \left(e^{2\pi i \frac{\nu_j}{d_j} y_j} - e^{2\pi i \frac{\nu_j}{d_j} x_j} \right).$$

In formulating our conjectures it is important to replace ω_f by a non-homolorphic differential ω_f^+ with the same associated Hecke eigenvalues. To do this let $\Sigma = \{\pm 1\}^n$ and for each $\sigma = (\sigma_1, \dots, \sigma_n) \in \Sigma$, let $\gamma_\sigma \in M_0(N)^\times$ be an element whose determinant is a unit of $\mathcal{O}_F^\times$ which is positive at ι_0 and the places ι_j for which $\sigma_j = 1$, and negative at the places ι_j for which $\sigma_j = -1$. (Such units exist, thanks to the narrow class number one assumption that was made for F.) For $\tau_j \in \mathcal{H}_j$ write

$$\tau_j^\sigma = \begin{cases} \gamma_\sigma \tau_j & \text{if } \sigma_j = 1 \text{ or } j = 0, \\ \gamma_\sigma \bar{\tau}_j & \text{if } \sigma_j = -1, \end{cases}$$

and set

$$\omega_f^\sigma = (2\pi i)^{n+1} f(\tau_0^\sigma, \tau_1^\sigma, \dots, \tau_n^\sigma) d\tau_0^\sigma d\tau_1^\sigma \cdots d\tau_n^\sigma.$$

Finally write:

$$\omega_f^+ := \sqrt{|d|}^{-1} \sum_{\sigma \in \Sigma} \omega_f^\sigma. \tag{8.2}$$

Properties of the multiple integral: The multiple integral of equation (8.1) enjoys the obvious additivity and Γ-invariance properties, such as

$$\left(\int_{x_0}^{y_0}\cdots\int_{x_j}^{t_j}\cdots\int_{x_n}^{y_n}\omega_f\right)+\left(\int_{x_0}^{y_0}\cdots\int_{t_j}^{y_j}\cdots\int_{x_n}^{y_n}\omega_f\right)=\int_{x_0}^{y_0}\cdots\int_{x_j}^{y_j}\cdots\int_{x_n}^{y_n}\omega_f, \tag{8.3}$$

for all $t_j \in \mathcal{H}_j$ and for all $j = 0, \ldots, n$. Also,

$$\int_{\gamma x_0}^{\gamma y_0}\cdots\int_{\gamma x_n}^{\gamma y_n}\omega_f=\int_{x_0}^{y_0}\cdots\int_{x_n}^{y_n}\omega_f, \quad \text{for all } \gamma\in\Gamma.$$

The same identities continue to hold with ω_f replaced by ω_f^+. The integral attached to ω_f^+ also enjoys an invariance property under the larger group $M_0(N)^\times$, which contains Γ with index 2^{n+1}. (Cf. exercise 1.) More precisely, extend the action of Γ on $\mathcal{H}_j$ to an action of $M_0(N)^\times$ by setting, for all $\tau \in \mathcal{H}_j$:

$$\gamma\tau=\begin{cases}\gamma_j\tau & \text{if } \det(\gamma_j)>0;\\ \gamma_j\bar{\tau} & \text{if } \det(\gamma_j)<0.\end{cases} \tag{8.4}$$

Let $c : \mathbb{C} \longrightarrow \mathbb{C}$ denote complex conjugation. For $\gamma \in M_0(N^\times)$, set $s_\gamma = 0$ (resp. $s_\gamma = 1$) if $\det(\gamma_0) > 0$ (resp. if $\det(\gamma_0) < 0$).

Lemma 8.1. *For all $x_j, y_j \in \mathcal{H}_j$, with $0 \leq j \leq n$, and all $\gamma \in M_0(N)^\times$,*

$$\int_{\gamma_0\bar{x}_0}^{\gamma_0\bar{y}_0}\int_{\gamma_1 x_1}^{\gamma_1 y_1}\cdots\int_{\gamma_n x_n}^{\gamma_n y_n}\omega_f^+=c^{s_\gamma}\left(\int_{x_0}^{y_0}\cdots\int_{x_n}^{y_n}\omega_f^+\right).$$

Proof. This is left to the reader as an exercise. (Cf. exercise 1.) □

Remark 8.2. Let $\tilde{\Gamma}$ be the group of elements $\gamma \in M_0(N)^\times$ for which $\det(\gamma_0)$ is positive. It follows from lemma 8.1 that the differential ω_f^+ is invariant under the action of $\tilde{\Gamma}$.

8.2. Generalities on group cohomology

We begin by briefly recalling some standard notations and terminology from group cohomology.

Let M be a $\mathbb{Z}$-module with trivial Γ-action. The space of M-valued r-cochains on Γ, denoted $C^r(\Gamma, M)$, is the set of functions $f : \Gamma^r \longrightarrow M$. Let

$$d : C^r(\Gamma, M) \longrightarrow C^{r+1}(\Gamma, M)$$

be the differential of degree r, defined as in [**CF67**], ch. 4:

$$\begin{aligned}(df)(g_1,\ldots,g_{r+1}) &= g_1 f(g_2,\ldots,g_{r+1})+\sum_{j=1}^{r}(-1)^j f(g_1,\ldots,g_jg_{j+1},\ldots,g_{r+1})\\ &\quad +(-1)^{r+1}f(g_1,\ldots,g_r).\end{aligned}$$

The space of M-valued r-cocycles on Γ, denoted $Z^r(\Gamma, M)$, is the submodule of $C^r(\Gamma, M)$ of functions f satisfying

$$df = 0.$$

The space of M-valued r-coboundaries, denoted $B^r(\Gamma, M)$, is the submodule of $C^r(\Gamma, M)$ consisting of functions of the form df, for some $f \in C^{r-1}(\Gamma, M)$. Since $dd = 0$, the module $B^r(\Gamma, M)$ is contained in $Z^r(\Gamma, M)$. The quotient module

$$H^r(\Gamma, M) = Z^r(\Gamma, M)/B^r(\Gamma, M)$$

is called the r-th cohomology group of Γ with coefficients in M.

An alternate definition of $H^r(\Gamma, M)$ in terms of *homogeneous cochains* will be useful for the calculations in this chapter. A *homogeneous r-cochain* is a function $f : \Gamma^{r+1} \longrightarrow M$ satisfying

$$f(sg_0, \ldots, sg_r) = sf(g_0, \ldots, g_r) \quad \text{for all } s \in \Gamma.$$

The space of homogeneous r-cochains is denoted $C^r_{\mathrm{hom}}(\Gamma, M)$: these groups form a complex in which the differential map $d : C^r_{\mathrm{hom}}(\Gamma, M) \longrightarrow C^{r+1}_{\mathrm{hom}}(\Gamma, M)$ is given by the simpler formula

$$(df)(g_0, \ldots, g_{r+1}) = \sum_{j=0}^{r+1} (-1)^j f(g_0, \ldots, g_{j-1}, g_{j+1}, \ldots, g_{r+1}).$$

One defines the group of homogeneous r-coboundaries $B^r_{\mathrm{hom}}(\Gamma, M)$ (resp. the group of homogeneous r-cocycles $Z^r_{\mathrm{hom}}(\Gamma, M)$) as the image of d (resp. the kernel of d).

Since a homogeneous r-cochain is determined by its values on r-tuples of the form $(1, g_1, g_1g_2, g_1g_2g_3, \ldots, g_1 \ldots g_r)$, one may associate to a homogeneous r-cochain f a non-homogeneous r-cochain θf by the rule

$$(\theta f)(g_1, \ldots, g_r) = f(1, g_1, g_1g_2, \ldots, g_1 \cdots g_r).$$

The commutativity of the diagram

$$\begin{array}{ccc} C^r_{\mathrm{hom}}(\Gamma, M) & \xrightarrow{d} & C^{r+1}_{\mathrm{hom}}(\Gamma, M) \\ \downarrow \theta & & \downarrow \theta \\ C^r(\Gamma, M) & \xrightarrow{d} & C^{r+1}(\Gamma, M) \end{array}$$

shows that the r-th cohomology of Γ with coefficients in M can also be computed as

$$H^r(\Gamma, M) = Z^r_{\mathrm{hom}}(\Gamma, M)/B^r_{\mathrm{hom}}(\Gamma, M).$$

8.3. The cohomology of Hilbert modular groups

Let τ be any element of $\mathcal{H}_0$. To the Γ-invariant differential form ω_f^+ and the point τ we associate a non-homogeneous $(n+1)$-cochain

$$\kappa_\tau \in C^{n+1}(\Gamma, \mathbb{C})$$

by choosing a base point $\underline{x} = (x_1, \ldots, x_n)$ in $\mathcal{H}_1 \times \cdots \times \mathcal{H}_n$, and setting

$$\kappa_\tau(\alpha_0, \alpha_1, \ldots, \alpha_n) = \int_\tau^{\alpha_0\tau} \int_{\alpha_0 x_1}^{\alpha_0\alpha_1 x_1} \int_{\alpha_0\alpha_1 x_2}^{\alpha_0\alpha_1\alpha_2 x_2} \cdots \int_{\alpha_0\cdots\alpha_{n-1}x_n}^{\alpha_0\cdots\alpha_n x_n} \omega_f^+. \tag{8.5}$$

LEMMA 8.3. *The $(n+1)$-cochain κ_τ is an $(n+1)$-cocycle, i.e., it belongs to $Z^{n+1}(\Gamma, \mathbb{C})$.*

PROOF. We show this by induction on n, the case $n = 1$ being checked by a direct computation. For the general case, let $\kappa'_\tau(g_0, \ldots, g_{n+1})$ be the homogeneous $(n+1)$-cochain attached to κ_τ, given by the formula:

$$\kappa'_\tau(g_0, \ldots, g_{n+1}) = \int_{g_0\tau}^{g_1\tau} \int_{g_1 x_1}^{g_2 x_1} \cdots \int_{g_n x_n}^{g_{n+1} x_n} \omega_f^+.$$

Fixing arbitray arguments g_{n+1} and $g_{n+2} \in \Gamma$, let $\kappa''_\tau : \Gamma^{n+1} \longrightarrow \mathbb{C}$ be the function given by

$$\kappa''_\tau(h_0, \ldots, h_n) = \int_{h_0\tau}^{h_1\tau} \int_{h_1x_1}^{h_2x_1} \int_{h_2x_2}^{h_3x_2} \cdots \int_{h_{n-1}x_{n-1}}^{h_nx_{n-1}} \int_{g_{n+1}x_n}^{g_{n+2}x_n} \omega_f^+.$$

The additivity property of the multiple integral combined with the induction hypothesis implies that

$$d\kappa''_\tau(h_0, \ldots, h_{n+1}) = 0, \quad \text{for all } h_0, \ldots, h_{n+1} \in \Gamma. \tag{8.6}$$

On the other hand, a direct calculation shows that

$$\begin{aligned} d\kappa'_\tau(g_0, \ldots, g_{n+2}) &= \sum_{j=0}^{n} (-1)^j \kappa'_\tau(g_0, \ldots, g_{j-1}, g_{j+1}, \ldots, g_{n+2}) \\ &+ (-1)^{n+1} \kappa'_\tau(g_0, \ldots, g_n, g_{n+2}) + (-1)^{n+2} \kappa'_\tau(g_0, \ldots, g_{n+1}) \end{aligned}$$

$$= d\kappa''_\tau(g_0, \ldots, g_{n+1}) + (-1)^n \left\{ \int_{g_0\tau}^{g_1\tau} \int_{g_1x_1}^{g_2x_1} \cdots \int_{g_{n-1}x_{n-1}}^{g_nx_{n-1}} \int_{g_{n+1}x_n}^{g_{n+2}x_n} \omega_f^+ \right.$$

$$\left. - \int_{g_0\tau}^{g_1\tau} \int_{g_1x_1}^{g_2x_1} \cdots \int_{g_{n-1}x_{n-1}}^{g_nx_{n-1}} \int_{g_nx_n}^{g_{n+2}x_n} \omega_f^+ + \int_{g_0\tau}^{g_1\tau} \int_{g_1x_1}^{g_2x_1} \cdots \int_{g_{n-1}x_{n-1}}^{g_nx_{n-1}} \int_{g_nx_n}^{g_{n+1}x_n} \omega_f^+ \right\}$$

But this last expression is equal to zero, by (8.6) and (8.3). □

In analysing the dependence of κ_τ on the choice of the base point $\underline{x}$, it is convenient to temporarily denote by $\kappa_{\tau,\underline{x}}$ the $(n+1)$-cocycle defined in equation (8.5). Let Γ_τ denote the stabiliser of τ in Γ.

LEMMA 8.4. *Let $\underline{x}$ and $\underline{y}$ be any two base points in $\mathcal{H}_1 \times \cdots \times \mathcal{H}_n$. There exists an n-cochain $\rho_{\underline{x},\underline{y}} \in C^n(\Gamma, \mathbb{C})$ such that*

(1) $\kappa_{\tau,\underline{x}} - \kappa_{\tau,\underline{y}} = d\rho_{\underline{x},\underline{y}}$,
(2) $\rho_{\underline{x},\underline{y}}(g_1, \ldots, g_n) = 0$, *for all* $g_1, \ldots, g_n \in \Gamma_\tau$.

PROOF. As in the previous proof it is easiest to see this by working with the homogeneous $n+1$-cocycles $\kappa'_{\tau,\underline{x}}$ and $\kappa'_{\tau,\underline{y}}$ attached to $\kappa_{\tau,\underline{x}}$ and $\kappa_{\tau,\underline{y}}$ respectively. Assume without loss of generality that $\underline{x}$ and $\underline{y}$ differ in a single coordinate, x_j say. Then the expression for $(\kappa'_{\tau,\underline{x}} - \kappa'_{\tau,\underline{y}})(g_0, \ldots, g_{n+1})$ is given by

$$\begin{aligned} (\kappa'_{\tau,\underline{x}} - \kappa'_{\tau,\underline{y}}) &= \int_{g_0\tau}^{g_1\tau} \cdots \int_{g_{j-1}x_{j-1}}^{g_jx_{j-1}} \int_{g_jx_j}^{g_jy_j} \int_{g_{j+1}x_{j+1}}^{g_{j+2}x_{j+1}} \cdots \int_{g_nx_n}^{g_{n+1}x_n} \omega_f^+ \\ &- \int_{g_0\tau}^{g_1\tau} \cdots \int_{g_{j-1}x_{j-1}}^{g_jx_{j-1}} \int_{g_{j+1}x_j}^{g_{j+1}y_j} \int_{g_{j+1}x_{j+1}}^{g_{j+2}x_{j+1}} \cdots \int_{g_nx_n}^{g_{n+1}x_n} \omega_f^+. \end{aligned}$$

A direct calculation (cf. exercise 2) reveals that this expression is equal to $d\rho_{\underline{x},\underline{y}}$, where $\rho_{\underline{x},\underline{y}} : \Gamma^{n+1} \longrightarrow \mathbb{C}$ is defined by

$$\rho_{\underline{x},\underline{y}}(g_0, \ldots, g_n) = \int_{g_0\tau}^{g_1\tau} \int_{g_1x_1}^{g_2x_1} \cdots \int_{g_{j-1}x_{j-1}}^{g_jx_{j-1}} \int_{g_jx_j}^{g_jy_j} \int_{g_jx_{j+1}}^{g_{j+1}x_{j+1}} \cdots \int_{g_{n-1}x_n}^{g_nx_n} \omega_f^+. \tag{8.7}$$

The result follows immediately from this formula. □

REMARK 8.5. Lemma 8.4 and its proof imply, in particular, that the natural image of κ_τ in $H^{n+1}(\Gamma, \mathbb{C})$ is an invariant of ω_f^+ which does not depend on the choice of base point $\underline{x}$, or, for that matter, of τ, that was made to define it.

A lattice $\Lambda \subset \mathbb{C}$ is called a *trivialising lattice* for the cocycle κ_τ if the natural image of κ_τ in $H^{n+1}(\Gamma, \mathbb{C}/\Lambda)$ is trivial. The key conjecture 8.6 below postulates the existence of a trivialising lattice for κ_τ which can be defined explicitly in terms of the periods attached to E. More precisely, choose a Néron differential ω_E on E/F and let Ω_j^+, $(j = 1, \ldots, n)$ denote the real period attached to the elliptic curve E_j and the differential $\iota_j(\omega_E)$. Let Λ_0 denote the period lattice attached to the differential $\iota_0(\omega_E)$ on E_0. Note that ω_E is only well-defined up to multiplication by a unit in F; hence the periods Ω_j^+ are defined up to multiplication by a non-zero element of $\iota_j(\mathcal{O}_F^\times)$. Since the norms of these elements are ± 1, the lattice

$$\Lambda := \Omega_1^+ \cdots \Omega_n^+ \Lambda_0$$

is independent of the choice of a Néron differential ω_E.

Notice also that Λ is homothetic to Λ_0. Let

$$\Phi_0 : \mathbb{C}/\Lambda \longrightarrow E_0(\mathbb{C}) \tag{8.8}$$

be the Weiserstrass uniformisation which is inverse to the map given by integration of the differential form $\Omega_1^+ \cdots \Omega_n^+ \iota_0(\omega_E)$.

CONJECTURE 8.6. *There exists an integer t_E (depending only on E, and not on τ) such that the cocycle $t_E\kappa_\tau$ has Λ as a trivialising lattice.*

Note that by remark 8.5, such an integer t_E, if it exists, can be chosen independently of τ.

REMARK 8.7. In the case $n = 0$, conjecture 8.6 merely asserts that an integer multiple of the cocycle κ_τ given by

$$\kappa_\tau(\gamma) = \int_\tau^{\gamma\tau} \omega_f$$

becomes trivial modulo Λ, the Néron latice of E. But this is well-known: for instance if E is the strong Weil curve attached to the classical modular curve parametrisation, one may take t_E to be the *Manin constant* attached to E, an integer which is conjecturally always equal to one.

REMARK 8.8. Conjecture 8.6 is related, in the case where $n = 1$, to work of Oda [**Od82**] on periods for Hilbert modular surfaces.

Let $\bar{\kappa}_\tau$ denote the natural image of $t_E\kappa_\tau$ in $Z^{n+1}(\Gamma, \mathbb{C}/\Lambda)$. Assuming conjecture 8.6, the cocycle $\bar{\kappa}_\tau$ is an $(n+1)$-coboundary, i.e., it belongs to $B^{n+1}(\Gamma, \mathbb{C}/\Lambda)$. Thus it is possible to write

$$\bar{\kappa}_\tau = d\tilde{\xi}_\tau, \quad \text{with} \quad \tilde{\xi}_\tau \in C^n(\Gamma, \mathbb{C}/\Lambda). \tag{8.9}$$

Note that equation (8.9) makes $\tilde{\xi}_\tau$ well-defined only up to elements of $Z^n(\Gamma, \mathbb{C}/\Lambda)$. The following theorem shows that the element $\tilde{\xi}_\tau$ – or at least, some multiple of it – is in fact well-defined up to n-coboundaries.

THEOREM 8.9. *Suppose that n is odd. Then $H^n(\Gamma, \mathbb{R}/\mathbb{Z})$ has finite exponent.*

PROOF. This follows from an extension due to Harder [**Ha75**] of a result of Matsushima and Shimura [**MS78**] valid in the case where Γ acts on $\mathcal{H}^{n+1}$ with compact quotient. More precisely, theorem 6.3 of chapter III, §6 of [**Frei90**] states that

$$H^n(\Gamma, \mathbb{R}) = 0.$$

Hence $H^n(\Gamma, \mathbb{R}/\mathbb{Z})$ injects (via the connecting homomorphism arising from the exact sequence $0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{R} \longrightarrow \mathbb{R}/\mathbb{Z} \longrightarrow 0$) into the kernel of the map $H^{n+1}(\Gamma, \mathbb{Z}) \longrightarrow H^{n+1}(\Gamma, \mathbb{R})$. This kernel consists of the torsion in $H^{n+1}(\Gamma, \mathbb{Z})$ which is finite (since it is finitely generated) and the result follows. □

Assume now that n is odd, and let e_Γ denote the exponent of $H^n(\Gamma, \mathbb{C}/\Lambda)$ which is finite by theorem 8.9. (This boundedness of the exponent is the single – but crucial – stage of the construction where it becomes necessary to assume $n \neq 0$.) The n-cochain

$$\xi_\tau := e_\Gamma \tilde{\xi}_\tau$$

is then well-defined, modulo

$$e_\Gamma Z^n(\Gamma, \mathbb{C}/\Lambda) \subset B^n(\Gamma, \mathbb{C}/\Lambda).$$

Then let $\theta_\tau \in C^n(\Gamma_\tau, \mathbb{C}/\Lambda)$ be the restriction of ξ_τ to ${\Gamma_\tau}^n$.

LEMMA 8.10. *The n-cochain θ_τ is an n-cocycle.*

PROOF. This follows from the fact, immediate from (8.5), that the $(n+1)$-cochain κ_τ vanishes identically on ${\Gamma_\tau}^{n+1}$. □

LEMMA 8.11. *The natural image of θ_τ in $H^n(\Gamma_\tau, \mathbb{C}/\Lambda)$ does not depend on the choice of $\tilde{\xi}_\tau$ made to define it.*

PROOF. This is because ξ_τ is well-defined up to n-coboundaries on Γ. □

LEMMA 8.12. *The natural image of θ_τ in $H^n(\Gamma_\tau, \mathbb{C}/\Lambda)$ does not depend on the choice of base points $\underline{x}$ that were made to define κ_τ and $\bar{\kappa}_\tau$.*

PROOF. This follows from lemma 8.4. (Cf. exercise 3.) □

Conclusion: To any point $\tau \in \mathcal{H}_0$ has been associated (at least when n is odd) a cohomology class

$$\theta_\tau \in H^n(\Gamma_\tau, \mathbb{C}/\Lambda).$$

This object depends only on the orbit of τ under the action of the group $\tilde{\Gamma}$ of remark 8.2. More precisely, suppose that $\tau \in \mathcal{H}_0$ and $\tau' = \alpha\tau$ belong to the same orbit for this group. Then $\Gamma_{\tau'} = \alpha\Gamma_\tau\alpha^{-1}$, and we have

LEMMA 8.13. *The classes of θ_τ and $\theta_{\tau'}$ admit representative cocycles $\tilde{\theta}_\tau$ and $\tilde{\theta}_{\tau'}$ (in $Z^n(\Gamma_\tau, \mathbb{C}/\Lambda)$ and $Z^n(\Gamma_{\tau'}, \mathbb{C}/\Lambda)$, respectively) such that*

$$\tilde{\theta}_{\tau'}(\alpha g_1 \alpha^{-1}, \ldots, \alpha g_n \alpha^{-1}) = \tilde{\theta}_\tau(g_1, \ldots, g_n),$$

for all $g_1, \ldots, g_n \in \Gamma_\tau$.

PROOF. See exercise 5. □

8.4. ATR points

Let K be a quadratic extension of F which is ATR (relative to the chosen ordering $v_0, \ldots, v_n$ of the archimedean places of F). Let $\tau \in \mathcal{H}_0 \cap K$ be an ATR point, so that the group Γ_τ is a free abelian group of rank n. In that case we have

PROPOSITION 8.14. *The cohomology group $H^n(\Gamma_\tau, \mathbb{C}/\Lambda)$ is canonically isomorphic to $\mathbb{C}/\Lambda$.*

PROOF. This follows directly from the fact that $\Gamma_\tau \simeq \mathbb{Z}^n$ can be made to act freely on $\mathbb{R}^n$ with quotient $T^n = R^n/\mathbb{Z}^n$, so that the cohomology of $\mathbb{Z}^n$ is identified with the cohomology of a compact connected oriented n-manifold. $\square$

Let $J_\tau \in \mathbb{C}/\Lambda$ be the invariant attached to θ_τ by proposition 8.14. It follows from lemma 8.13 that

$$J_{\alpha\tau} = J_\tau, \quad \text{for all } \alpha \in \Gamma.$$

Conclusion: To any ATR point $\tau \in \mathcal{H}_0$ has been associated a canonical invariant $J_\tau \in \mathbb{C}/\Lambda$, which depends only on the $\tilde{\Gamma}$-orbit of $\tau \in \mathcal{H}_0$. Recall the Weierstrass uniformisation Φ_0 of equation (8.8) and set

$$P_\tau := \Phi_0(J_\tau) \in E_0(\mathbb{C}).$$

We can now define the map Φ'_N alluded to at the end of chapter 7 by the rule

$$\Phi'_N(\tau) = P_\tau.$$

We now formulate a conjecture, analogous to the classical Shimura reciprocity law of chapter 3, which implies that the point $\Phi'_N(\tau)$ (with $\tau \in \mathcal{H}_0 \cap K$ an ATR point) is an algebraic point on E defined over an abelian extension of K which can be described precisely in terms of class field theory.

The following assumption, while not essential, will be made to simplify the discussion, and is analogous to the Heegner hypothesis of chapter 3:

HYPOTHESIS 8.15. *All the primes dividing the conductor N of E are split in K/F.*

Fix an ideal $\mathcal{N}$ of $\mathcal{O}_K$ whose norm to $\mathcal{O}_F$ is equal to N. Recall that $M_0(N)$ is the $\mathcal{O}_F$-algebra of 2×2 matrices with entries in $\mathcal{O}_F$ which are upper-triangular modulo N. Denote by

$$\eta : M_0(N) \longrightarrow \mathcal{O}_F/N\mathcal{O}_F$$

the homomorphism sending a matrix to its upper left-hand entry taken modulo N. Let

$$\Psi : K \longrightarrow M_2(F)$$

be an embedding of F-algebras. Such an embedding is said to be *optimal* if

$$\Psi(K) \cap M_0(N) = \Psi(\mathcal{O}_K).$$

Let $\tau \in \mathcal{H}_0$ denote the unique fixed point for $\Psi(K^\times)$ acting on $\mathcal{H}_0$. The embedding Ψ is said to be *normalised* if $\Psi(\alpha)$ acts on the column vector $(\tau, 1)$ as multiplication by α, and if the kernel of $\eta\Psi$ (restricted to $\mathcal{O}_K$) is equal to $\mathcal{N}$.

The group $\tilde{\Gamma}$ acts naturally by conjugation on the set of normalised optimal embeddings.

Let H denote the Hilbert class field of K, and let G denote its Galois group over K. This group G is identified with the Picard group of $\mathcal{O}_K$.

The group G acts naturally on the collection of $\tilde{\Gamma}$-conjugacy classes of normalised optimal embeddings of K into $M_2(F)$ (cf. exercise 6). Given $\sigma \in G$, and a normalised embedding Ψ, let $\sigma \star \Psi$ denote a representative for the $\tilde{\Gamma}$-orbit of the image of the embedding Ψ by the action of σ.

LEMMA 8.16. *The group G acts simply transitively on the set of $\tilde{\Gamma}$-conjugacy classes of normalised optimal embeddings of K into $M_2(F)$.*

PROOF. See exercise 6. $\square$

Given a normalised embedding Ψ with fixed point $\tau \in \mathcal{H}_0$, one may set

$$P_\Psi = \Phi'_N(\tau).$$

Note that this point depends only on the $\tilde{\Gamma}$-conjugacy class of Ψ.

We view the Hilbert class field H of K as a subfield of $\mathbb{C}$ via a complex embedding which extends the fixed embedding v_0 of K into $\mathbb{C}$.

CONJECTURE 8.17. *The local point P_Ψ is the image of a global point in $E(H)$. More precisely*

$$P_{\sigma\star\Psi} = \sigma^{-1}(P_\Psi), \qquad \textit{for all } \sigma \in H.$$

REMARK 8.18. The simplest case of the construction (and the one which runs the most parallel to the p-adic construction of chapter 9) is the case where $n = 1$. The field F is then a real quadratic field and K is a quadratic extension of F having one complex and two real places. The space X is a Hilbert modular surface attached to the Hecke congruence subgroup of level N in $\mathbf{SL}_2(\mathcal{O}_F)$.

REMARK 8.19. Specialising even further, suppose that F is a real quadratic field of narrow class number one, and that E is an elliptic curve with everywhere good reduction over F, so that $N = 1$. Let K be a quadratic extension of F with one complex and two real places. In this situation conjecture 8.17 yields an analytic construction of points on $E(H)$, where H is the Hilbert class field of K. This setting is of interest because (1) if E is not isogenous to its Galois conjugate then E does not appear as a quotient of the Jacobian of any modular or Shimura curve, so that the theory of complex multiplication does not suggest any modular construction of Heegner systems on E; and (2) this situation is simple enough to be amenable to machine calculations. For some numerical verifications that have been performed in this case, see [**DL03**].

REMARK 8.20. The only ATR extensions of $F = \mathbb{Q}$ are the imaginary quadratic fields, where one is placed in the usual setting of the theory of complex multiplication. There are "not enough" archimedean places of $\mathbb{Q}$ for any new phenomena to manifest themselves through conjecture 8.17. In particular, since a real quadratic field is not an ATR extension, the generalisation of the Heegner point construction to the context of real quadratic fields falls outside the scope of the theory developed in the last two chapters. To encompass this case it is necessary to combine this theory with some of the notions arising from rigid analysis that were surveyed in chapters 5 and 6. This will be the goal of the next chapter.

References

A reference for periods of Hilbert modular forms (at least in the case of Hilbert modular surfaces) is the book of Oda [**Od82**].

For an introduction to the cohomology of groups, and in particular to the cohomology of discrete subgroups of Lie groups, see [**Br94**]. General finiteness results on the cohomology of discrete groups (particularly arithmetic groups) can be found in the articles [**Se70**] and [**Se71**] by Serre. The cohomology of Hilbert modular groups, building on the techniques of Matsushima-Shimura [**MS78**] and Harder [**Ha75**] is studied in [**Frei90**].

For further discussion of conjecture 8.17, see [**BDG03**] and [**DL03**]. The article [**DL03**] presents numerical evidence for conjecture 8.17 when E has everywhere good reduction over a real quadratic field, and discusses the algorithms that were used to calculate Φ'_N in this case.

Exercises

(1) Assuming that F has narrow class number one, show that Γ is a normal subgroup of $M_0(N)^\times$ with quotient isomorphic to $(\mathbb{Z}/2\mathbb{Z})^{n+1}$. Write down representatives for the cosets of Γ in $M_0(N)^\times$. (Hint: they can be chosen to be diagonal matrices.) Use these representatives to prove lemma 8.1.

(2) Complete the proof of lemma 8.4 by showing that the homogeneous n-cochain $\rho_{\underline{x},\underline{y}}$ of equation 8.7 satisfies $d\rho = \kappa_{\tau,\underline{x}} - \kappa_{\tau,\underline{y}}$.

(3) Provide the details of the proof of lemma 8.12.

(4) Let $G = \mathbb{Z}^n$ and let M be a G-module with trivial G-action. Show that the isomorphism between $H^n(G, M)$ and M can be written in terms of non-homogeneous n-cochains by choosing a system of generators $\gamma_1, \ldots, \gamma_n$ of G and sending the class of the cocycle θ to the quantity
$$J_\theta := \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma)\theta_\tau(\gamma_{\sigma 1}, \gamma_{\sigma 2}, \ldots, \gamma_{\sigma n}),$$
where $\operatorname{sgn}(\sigma) = \pm 1$ is the sign of the permutation σ.

(5) Prove lemma 8.13, and conclude that the invariant J_τ attached to an ATR point $\tau \in \mathcal{H}_0$ depends only on the $\tilde{\Gamma}$-orbit of τ, so that Φ'_N descends to a map
$$\Phi'_N : \mathcal{H}'_0/\tilde{\Gamma} \longrightarrow E(\mathbb{C}).$$

(6) Let K be a fixed ATR extension of F in which all the prime divisors of N are split. Adapt the ideas of sections 3.1 and 3.2 to define a natural action of $G = \operatorname{Pic}(\mathcal{O}_K)$ on the collection of $\tilde{\Gamma}$- conjugacy classes of normalised optimal embeddings of K into $M_2(F)$ (relative to $M_0(N)$, and a choice of ideal $\mathcal{N} \subset \mathcal{O}_K$). Show that this action is simply transitive.

CHAPTER 9

Integration on $\mathcal{H}_p \times \mathcal{H}$

A number of different modular parametrisations have been introduced so far, and their attendant theories of special (CM, or ATR) points used to construct Heegner systems attached to modular elliptic curves.

In **chapters 2 and 3**, the classical modular parametrisation

$$\Phi_N : \mathcal{H}/\Gamma_0(N) \longrightarrow E(\mathbb{C})$$

attached to a modular elliptic curve over $\mathbb{Q}$ of conductor N, combined with the theory of complex multiplication, was used to define global points on E over ring class fields of quadratic imaginary fields satisfying a suitable Heegner hypothesis.

In **chapters 4 and 7**, the Shimura curve parametrisations

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

were used to construct more general Heegner systems attached to elliptic curves over $\mathbb{Q}$ and quadratic imaginary extensions of $\mathbb{Q}$. When suitably generalised to groups coming from quaternion algebras over a totally real field F, Shimura curve parametrisations may be used to construct Heegner systems attached to elliptic curves over F and certain CM extensions of F.

In **chapter 6**, the rigid analytic uniformisation

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

based on the Čerednik-Drinfeld theory of p-adic uniformisation of Shimura curves was then exploited to construct the Heegner systems of chapter 4 by p-adic analytic means.

Finally, when E is a modular elliptic curve of conductor $\mathfrak{n}$ defined over a totally real field F of degree $n+1 > 1$, and f is the associated Hilbert modular form on $(\mathcal{H}_0 \times \cdots \times \mathcal{H}_n)/\Gamma_0(\mathfrak{n})$, modular or Shimura curve parametrisations are not always available. **Chapters 7 and 8** introduced a *natural substitute*

$$\Phi'_N : \mathcal{H}'_0/\Gamma_0(\mathfrak{n}) \longrightarrow E(\mathbb{C})$$

defined on the set $\mathcal{H}'_0$ of ATR points in $\mathcal{H}_0$ in terms of certain integrals attached to f. This substitute can be used – conjecturally – to construct points on elliptic curves defined over abelian extensions of ATR extensions of F.

This chapter proposes a synthesis of the ideas of chapters 6 and 8 to define (*conjecturally*, as in chapter 8) Heegner systems attached to pairs (E, K) where E is a modular elliptic curve over $\mathbb{Q}$ and K is a real quadratic field for which $\mathrm{sign}(E, K) = -1$.

Note that, if $\mathrm{sign}(E,K) = -1$, then there is at least one prime divisor p of N which is either inert or ramified in $K/\mathbb{Q}$, by theorem 3.17 of chapter 3. The main idea of the construction is to fix such a prime divisor and make it play the same role as the distinguished archimedean place v_0 in the theory of the two previous chapters.

9.1. Discrete arithmetic subgroups of $\mathbf{SL}_2(\mathbb{Q}_p) \times \mathbf{SL}_2(\mathbb{R})$

Suppose henceforth that the conductor N of E is of the form pM, with p a prime not dividing M. In this case E has multiplicative reduction at p by (1.17) of chapter 1. Let q be the Tate period attached to $E/\mathbb{Q}_p$ and let

$$\Phi_{\mathrm{Tate}} : \mathbb{C}_p^\times \longrightarrow E(\mathbb{C}_p)$$

be the Tate uniformisation attached to E as in equation (1.9) of chapter 1. Set $w = a_p$, so that

$$w = \begin{cases} 1 & \text{if } E \text{ has split multiplicative reduction at } p. \\ -1 & \text{if } E \text{ has non-split multiplicative reduction at } p. \end{cases} \tag{9.1}$$

Let $\Gamma \subset \mathbf{SL}_2(\mathbb{Z}[1/p])$ be the group of matrices which are upper triangular modulo M. This group acts naturally both on $\mathcal{H}$ and on the p-adic upper half plane $\mathcal{H}_p$ of chapter 5. It acts on each of these upper half planes with dense orbits, but its action on the product $\mathcal{H}_p \times \mathcal{H}$ is discrete. (Cf. exercise 1.) Note that Γ is the group of elements of determinant one in the ring $R = M_0(M)[1/p]$ of matrices with entries in $\mathbb{Z}[1/p]$ which are upper-triangular modulo M. Let $\tilde{\Gamma} = R^\times$ and let α_∞ and α_p be elements of $\tilde{\Gamma}$ satisfying

$$\det(\alpha_\infty) = -1, \quad \det(\alpha_p) = p. \tag{9.2}$$

A point τ belonging to $\mathcal{H}$ or $\mathcal{H}_p$ is called a *special point* if its stabiliser in Γ is infinite. Let $\mathcal{H}'$ and $\mathcal{H}_p'$ denote the set of special points in $\mathcal{H}$ and $\mathcal{H}_p$ respectively. It can be seen that $\mathcal{H}'$ consists precisely of the points in $\mathcal{H} \cap K$, where K ranges over all the imaginary quadratic extensions of $\mathbb{Q}$ in which p is split. In [**Ih68**], Ihara describes a natural correspondence between $\mathcal{H}'$ and certain points on $X_0(M)$ in characteristic p, this correspondence being derived via the classical theory of complex multiplication developed in chapter 3. (Cf. exercise 2.)

On the other hand, the set $\mathcal{H}_p'$ is an object that appears to have been less studied. One can write

$$\mathcal{H}_p' = \cup_K (\mathcal{H}_p \cap K),$$

where K ranges over all real quadratic subfields of $\mathbb{C}_p$ which are not contained in $\mathbb{Q}_p$ (so that p is inert or ramified in K).

The goal of this chapter is to explain in detail how the following program can be carried out:

(1) Attach to the elliptic curve E a "mock Hilbert modular form of weight $(2,2)$" on $(\mathcal{H}_p \times \mathcal{H})/\Gamma$, denoted ω.
(2) Define a $\mathbb{C}_p$-valued integration theory attached to ω, behaving formally like the integrals attached to a Hilbert modular form in chapter 8. The ideas of p-adic analysis and p-adic integration explained in chapter 5, together with the Manin-Drinfeld theorem described in section 2.7, are the main ingredients in this definition.

(3) Use these periods to define a parametrisation

$$\Phi_N^{\prime(p)} : \mathcal{H}_p'/\Gamma \longrightarrow E(\mathbb{C}_p)$$

by mimicking the formal aspects of the definition of the parametrisation Φ_N' given in chapter 8;

(4) Make a precise conjecture predicting that the points $\Phi_N^{\prime(p)}(\tau) \in E(\mathbb{C}_p)$ are defined over certain ring class fields of real quadratic fields.

9.2. Forms on $\mathcal{H}_p \times \mathcal{H}$

Motivated by the analogy with Hilbert modular forms, we seek an appropriate notion of "form of weight $(2,2)$ on $(\mathcal{H}_p \times \mathcal{H})/\Gamma$". In an informal sense, such an object would be a Γ-invariant expression of the form

$$\text{“}\omega = f(z_p, z)dz_p dz\text{”},$$

where z_p is a p-adic and z is a complex variable. If one requires that the function f be rigid analytic in the first variable and holomorphic in the second, one runs into difficulties in making such a notion mathematically precise. While a sensible definition of ω seems elusive, the p-adic boundary distribution that was attached to a rigid analytic modular form in chapter 5 is a notion that can be transposed in a natural way to the current setting.

Recall the Bruhat-Tits tree $\mathcal{T}$ of $\mathbf{PGL}_2(\mathbb{Q}_p)$, whose set $\mathcal{T}_0$ of vertices is in bijection with the set of $\mathbb{Q}_p^\times$-homothety classes of rank two $\mathbb{Z}_p$-modules in $\mathbb{Q}_p^2$. Recall that $\mathcal{T}_1$ and $\mathcal{E}(\mathcal{T})$ are the set of unordered and ordered edges of $\mathcal{T}$ respectively, and that $s(e)$ and $t(e)$ denote the source and target vertex of $e \in \mathcal{E}(\mathcal{T})$. Let $\bar{e}$ denote the oriented edge obtained from e by interchanging its source and target.

The following definition, motivated by the considerations of chapter 5 and 6, is tailored to capture the notion of the "p-adic boundary distribution attached to ω".

Definition 9.1. A cusp form of weight 2 on $(\mathcal{T} \times \mathcal{H})/\Gamma$ is a function

$$f : \mathcal{E}(\mathcal{T}) \times \mathcal{H} \longrightarrow \mathbb{C}$$

satisfying

(1) $f(\gamma e, \gamma z) = (cz+d)^2 f(e,z)$, for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$.

(2) The function f is *harmonic*, i.e., for each vertex v of $\mathcal{T}$,

$$\sum_{s(e)=v} f(e,z) = 0,$$

and for each oriented edge e of $\mathcal{T}$, $f(\bar{e}, z) = -f(e,z)$.

(3) For each oriented edge e of $\mathcal{T}$, the function $f_e(z) := f(e,z)$ is a cusp form of weight 2 (in the usual sense) on the group $\Gamma_e := \mathrm{Stab}_\Gamma(e)$.

Property 1 is suggested by the desired Γ-invariance of ω, and property 2 by the ideas in the proof of theorem 5.9 of chapter 5. Note that an element of the space $S_2(\mathcal{T}, \Gamma)$ of cusp forms of weight 2 on $(\mathcal{T} \times \mathcal{H})/\Gamma$ is completely described by a collection $\{f_e\}$ of cusp forms on Γ_e, indexed by the ordered edges e of $\mathcal{T}$, satisfying the compatibility relation

$$f_{\gamma e}(\gamma z) d(\gamma z) = f_e(z) dz, \text{ for all } \gamma \in \Gamma, \tag{9.3}$$

together with the harmonicity condition 2.

Let e_o be the distinguished oriented edge of $\mathcal{T}$ whose stabiliser in Γ is equal to $\Gamma_0(N)$, and whose source is fixed by $\Gamma_0(M)$. The group Γ (resp. $\tilde{\Gamma}$) acts transitively on the unoriented (resp. oriented) edges of $\mathcal{T}$ (cf. exercise 1), and for each e the stabiliser subgroups Γ_e are conjugate in Γ to either $\Gamma_{e_o} = \Gamma_0(N)$ or $\Gamma_{\bar{e}_0}$.

The matrices α_∞ and α_p of equation (9.2) generate the quotient $\tilde{\Gamma}/\Gamma$, and can be chosen in such a way that they preserve the unordered edge attached to e_o, so that

$$\alpha_\infty e_o = e_o, \quad \alpha_p e_o = \bar{e}_o. \tag{9.4}$$

The involution W_p on the space $S_2(\mathcal{T}, \Gamma)$ is defined by the rule

$$(W_p f)(e, z)dz = f(\alpha_p e, \alpha_p z)d(\alpha_p z).$$

This definition does not depend on the choice of α_p that was made. If α_p is in addition normalised to satisfy (9.4) (a convention that will be adopted from now on) then it can also be used to define the usual Atkin-Lehner involution on the space $S_2(N) = S_2(\Gamma_0(N))$ of classical cusp forms of level N (denoted W_p as well by an abuse of notation) by the rule

$$(W_p f_0)(z)dz = f_0(\alpha_p z)d(\alpha_p z) \quad \text{for } f_0 \in S_2(N). \tag{9.5}$$

Lemma 9.2 below reveals that the space $S_2(\mathcal{T}, \Gamma)$ is intimately connected with $S_2(N)$. In order to state it precisely, let

$$\varphi_s : X_0(N) \longrightarrow X_0(M)$$

be the projection arising from the inclusion $\Gamma_0(N) \subset \Gamma_0(M)$, and let $\varphi_t = \varphi_s W_p$ Making an abuse of notation, denote by the same symbols φ_s and φ_t the two *degeneracy maps* from $S_2(N)$ to $S_2(M)$ induced from φ_s and φ_t by pushforward of differential forms. More precisely, choosing a system of coset representatives for $\Gamma_0(N)$ in $\Gamma_0(M)$:

$$\Gamma_0(M) = \gamma_1\Gamma_0(N) \cup \cdots \cup \gamma_{p+1}\Gamma_0(N), \tag{9.6}$$

one has

$$\varphi_s(f)(z)dz = \sum_{j=1}^{p+1} f(\gamma_j^{-1}z)d(\gamma_j^{-1}z), \quad \varphi_t(f)(z)dz = \sum_{j=1}^{p+1} f(\alpha\gamma_j^{-1}z)d(\alpha\gamma_j^{-1}z). \tag{9.7}$$

The kernel of

$$\varphi_s \oplus \varphi_t : S_2(N) \longrightarrow S_2(M) \oplus S_2(M) \tag{9.8}$$

is called the subspace of *p-new* forms, denoted $S_2^{\mathrm{new-p}}(N)$.

LEMMA 9.2. *The function which to $f(e,z)$ associates $f_o(z) := f_{e_o}(z)$ induces an isomorphism from $S_2(\mathcal{T}, \Gamma)$ to $S_2^{\mathrm{new-p}}(N)$.*

PROOF. See exercise 3 or the proof of lemma 1.3 of [**Da01**]. □

The Hecke operators T_ℓ ($\ell \nmid N$) act on the space $S_2(\mathcal{T}, \Gamma)$ via the identifications of lemma 9.2. One can also give a direct description in the spirit of equations (2.6) of chapter 2 and (4.3) of chapter 4. For each prime ℓ, write the double coset $\Gamma \begin{pmatrix} \ell & 0 \\ 0 & 1 \end{pmatrix} \Gamma$ as a disjoint union of left cosets:

$$\Gamma \begin{pmatrix} \ell & 0 \\ 0 & 1 \end{pmatrix} \Gamma = \cup_j \gamma_j \Gamma. \tag{9.9}$$

Then $T_\ell f$ is given by

$$(9.10) \qquad (T_\ell f)(e,z)dz = \sum_j f(\gamma_j^{-1}e, \gamma_j^{-1}z)d(\gamma_j^{-1}z).$$

Let f_0 be the normalized newform on $\Gamma_0(N)$ associated to E by theorem 2.11 of chapter 2. The form f_0 is an eigenvector for the involution W_p acting on on $S_2(N)$. It is known that

$$(9.11) \qquad W_p f_0 = -w f_0,$$

where w is the sign that was attached to E in (9.1). Let f be the form in $S_2(\mathcal{T}, \Gamma)$ which is related to f_0 by lemma 9.2, so that $f_{e_o} = f_0$. Given $\gamma \in \tilde{\Gamma}$, let $|\gamma|_p = \mathrm{ord}_p(\det(\gamma)) \in \mathbb{Z}/2\mathbb{Z}$. Write $R_+^\times$ for the group of elements in Γ of positive determinant.

LEMMA 9.3. *The form f satisfies the following transformation rule under the group $R_+^\times \supset \Gamma$:*

$$f(\gamma e, \gamma z)d(\gamma z) = w^{|\gamma|_p} f(e,z)dz, \qquad \textit{for all } \gamma \in R_+^\times.$$

PROOF. The lemma is clearly true for $\gamma \in \Gamma$. Since α_p generates the quotient $R_+^\times/\Gamma$, and f is determined by its behaviour on $\{e_o\} \times \mathcal{H}$, it suffices to check the lemma for $\gamma = \alpha_p$ and $e = e_o$. In that case one has:

$$\begin{aligned} f(\alpha_p e_0, \alpha_p z)d(\alpha_p z) &= f(\bar{e}_0, \alpha_p z)d(\alpha_p z) = -(W_p f_0)(z)dz \\ &= w f_0(z)dz = w f(e_o, z)dz, \end{aligned}$$

where the penultimate equality follows from (9.11). □

9.3. Periods

It will be convenient (but not essential) to make the following simplifying assumption:

ASSUMPTION 9.4. *The elliptic curve E is unique in its $\mathbb{Q}$-isogeny class.*

By this assumption, E is isomorphic to the strong Weil curve in its isogeny class. Let

$$\Phi_N : \mathcal{H}^*/\Gamma_0(N) \longrightarrow E(\mathbb{C})$$

be the strong Weil parametrisation attached to E. Letting ω_E denote the Néron differential of E, one has

$$(9.12) \qquad \varphi^*(\omega_E) = c \cdot 2\pi i f_0(z)dz,$$

where c is the *Manin constant*, introduced in (2.16) of section 2.5.

Choose elements x, y in the extended upper half-plane $\mathcal{H}^* := \mathcal{H} \cup \mathbb{P}_1(\mathbb{Q})$. The function $\tilde{\kappa}_f\{x\to y\} : \mathcal{E}(\mathcal{T}) \longrightarrow \mathbb{C}$ defined by

$$(9.13) \qquad \tilde{\kappa}_f\{x\to y\}(e) := c \cdot 2\pi i \int_x^y f_e(z)dz$$

is a complex-valued harmonic cocycle on $\mathcal{T}$, as follows immediately from the harmonicity properties of f itself. As in equation (5.5) of section 5.2, the harmonic cocycle $\tilde{\kappa}_f\{x\to y\}$ gives rise to a complex-valued *distribution* $\tilde{\mu}_f\{x\to y\}$ on the boundary $\mathbb{P}_1(\mathbb{Q}_p)$ of $\mathcal{H}_p$ by the rule

$$(9.14) \qquad \tilde{\mu}_f\{x\to y\}(U_e) = \tilde{\kappa}_f\{x\to y\}(e).$$

(Here $U_e \subset \mathbb{P}_1(\mathbb{Q}_p)$ is the compact open subset attached to $e \in \mathcal{E}(\mathcal{T})$ as in chapter 5.) Using equation (5.3) of chapter 5, the distribution $\tilde{\mu}_f\{x{\to}y\}$ can be integrated against locally constant complex-valued functions on $\mathbb{P}_1(\mathbb{Q}_p)$. For the purposes of p-adic integration, it is desirable that $\tilde{\kappa}_f\{x{\to}y\}$ take on integral or at least p-adic integral values and thereby correspond to a *measure* against which locally analytic functions can be integrated.

Such an integrality can be achieved by invoking the theory of modular symbols presented in section 2.7, *provided* that x and y belong to $\mathbb{P}_1(\mathbb{Q})$. For in this case, the value of $\tilde{\kappa}_f\{x{\to}y\}(e)$ can be expressed in terms of the modular symbols

$$\tilde{\lambda}_{f_0}\{x{\to}y\} := c \cdot 2\pi i \int_x^y f_0(z)dz. \tag{9.15}$$

(Here $\tilde{\lambda}_{f_0}$ is a slight modification of the M-symbol denoted λ_{f_0} in equation (2.18) of section 2.7.) More precisely, choosing $\gamma \in R_+^\times$ such that $\gamma e = e_o$,

$$\tilde{\kappa}_f\{x{\to}y\}(e) = w^{|\gamma|_p} \tilde{\lambda}_{f_0}\{\gamma x{\to}\gamma y\}. \tag{9.16}$$

Let Λ_E denote the Néron lattice attached to the elliptic curve E. Theorem 2.19 of section 2.7 combined with assumption 9.4 implies the following corollary:

COROLLARY 9.5. *For all $x, y \in \mathbb{P}_1(\mathbb{Q})$, the harmonic cocycle $\tilde{\kappa}_f\{x{\to}y\}$ takes its values in Λ_E, and hence gives rise to a Λ_E-valued measure $\mu_f\{x{\to}y\}$ on $\mathbb{P}_1(\mathbb{Q}_p)$.*

If $E(\mathbb{R})$ has two components then Λ_E is generated by a positive real period Ω_+ and a purely imaginary period Ω_-. If $E(\mathbb{R})$ has one connected component, then Ω is contained with index two in the lattice spanned by Ω_+ and Ω_-, where Ω_+ (resp. Ω_-) denotes the real (resp. imaginary) half-period attached to E. In either case, thanks to corollary 9.5, one can write

$$\tilde{\kappa}_f\{x{\to}y\}(e) = \kappa_f^+\{x{\to}y\}(e) \cdot \Omega_+ + \kappa_f^-\{x{\to}y\}(e) \cdot \Omega_-, \tag{9.17}$$

with $\kappa_f^\pm\{x{\to}y\}(e) \in \mathbb{Z}$. Choose a sign $w_\infty = \pm 1$ and let $\kappa_f\{x{\to}y\}$ denote $\kappa_f^+\{x{\to}y\}$ (resp. $\kappa_f^-\{x{\to}y\}$) if $w_\infty = 1$ (resp. $w_\infty = -1$). Write λ_{f_0} for the corresponding $\mathbb{Z}$-valued modular symbol attached to f_0. Let $\mu_f\{x{\to}y\}$ denote the $\mathbb{Z}$-valued distribution on $\mathbb{P}_1(\mathbb{Q}_p))$ attached to $\kappa_f\{x{\to}y\}$, so that, with the notations of (9.16),

$$\mu_f\{x{\to}y\}(U_e) := \kappa_f\{x{\to}y\}(e) = w^{|\gamma|_p} \lambda_{f_0}\{\gamma x{\to}\gamma y\}. \tag{9.18}$$

Given $\gamma \in \tilde{\Gamma}$, set

$$|\gamma|_\infty = \begin{cases} 0 & \text{if } \det(\gamma) > 0; \\ 1 & \text{if } \det(\gamma) < 0. \end{cases}$$

It is worth recording the following lemma which describes the behaviour of κ_f under the full group $\tilde{\Gamma}$.

LEMMA 9.6. *For all $\gamma \in \tilde{\Gamma}$, $x, y \in \mathbb{P}_1(\mathbb{Q})$, and $e \in \mathcal{E}(\mathcal{T})$,*

$$\kappa_f\{\gamma x{\to}\gamma y\}(\gamma e) = w_\infty^{|\gamma|_\infty} w^{|\gamma|_p} \kappa_f\{x{\to}y\}(e).$$

PROOF. If γ belongs to $R_+^\times$, this follows directly from the transformation properties of f given in lemma 9.3. It therefore suffices to prove the lemma for a single γ which generates $\tilde{\Gamma}/R_+^\times$, say the matrix $\gamma \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and for $e = e_o$. Then we have

$$\tilde{\kappa}_f\{\gamma x \to \gamma y\}(\gamma e_o) = \int_{-x}^{-y} f_0(z)dz = \tilde{\lambda}_{f_0}\{-x \to -y\} = \overline{\tilde{\lambda}_{f_0}\{x \to y\}}, \tag{9.19}$$

where the last equality follows from (2.22) and (2.23) of section 2.7. The lemma follows at once, given the definition of λ_f in terms of $\tilde{\lambda}_f$ and w_∞. ☐

Because the values $\mu_f\{x \to y\}(U_e)$ are integral and hence p-adically bounded as e varies in $\mathcal{E}(\mathcal{T})$, the distribution $\mu_f\{x \to y\}$ defines a p-adic *measure* on $\mathbb{P}_1(\mathbb{Q}_p)$. In particular, if h is any bounded locally analytic $\mathbb{C}_p$-valued function on $\mathbb{P}_1(\mathbb{Q}_p)$, the integral

$$\int_{\mathbb{P}_1(\mathbb{Q}_p)} h(t) d\mu_f\{x \to y\}(t) \in \mathbb{C}_p \tag{9.20}$$

can be defined as in equation (5.4) of chapter 5.

Inspired by definition 5.12 of chapter 5, the following definition, depending similarly on a choice of log, imposes itself naturally.

DEFINITION 9.7. Let z_1 and z_2 be elements of $\mathcal{H}_p$, and let $x, y \in \mathbb{P}_1(\mathbb{Q})$.

$$\int_{z_1}^{z_2}\int_x^y \omega := \int_{\mathbb{P}_1(\mathbb{Q}_p)} \log\left(\frac{t-z_2}{t-z_1}\right) d\mu_f\{x \to y\}(t) \in \mathbb{C}_p. \tag{9.21}$$

The following lemma shows that this definition is well-behaved:

LEMMA 9.8. *The double integrals of definition 9.7 satisfy the following properties:*

$$\int_{z_1}^{z_3}\int_x^y \omega = \int_{z_1}^{z_2}\int_x^y \omega + \int_{z_2}^{z_3}\int_x^y \omega, \quad \textit{for all } z_1, z_2, z_3 \in \mathcal{H}_p; \tag{9.22}$$

$$\int_{z_1}^{z_2}\int_{x_1}^{x_3} \omega = \int_{z_1}^{z_2}\int_{x_1}^{x_2} \omega + \int_{z_1}^{z_2}\int_{x_2}^{x_3} \omega, \quad \textit{for all } x_1, x_2, x_3 \in \mathbb{P}_1(\mathbb{Q}); \tag{9.23}$$

$$\int_{\gamma z_1}^{\gamma z_2}\int_{\gamma x}^{\gamma y} \omega = w^{|\gamma|_p} w_\infty^{|\gamma|_\infty} \int_{z_1}^{z_2}\int_x^y \omega, \quad \text{for all } \gamma \in R^\times. \tag{9.24}$$

PROOF. The first and second identity are a direct consequence of definition 9.7, while the third follows from lemma 9.6. See exercise 4 for details. ☐

We emphasize once again that the symbol ω is used as a placeholder and does not refer to an independently defined mathematical object. Only the form f in $S_2(\mathcal{T}, \Gamma)$ is defined, but this is enough to make sense of definition 9.7. The notation suggests that the left hand side of definition 9.7 be viewed as a period for a form of weight $(2, 2)$ on $(\mathcal{H}_p \times \mathcal{H})/\Gamma$, with the complex period Ω_+ or Ω_- "factored out".

To obtain stronger formulae, it is preferable to avoid choosing a p-adic logarithm, exploiting the fact that $\kappa_f\{x \to y\}$ is $\mathbb{Z}$-valued to make the same multiplicative refinement as in chapter 5 to define

$$⨰_{z_1}^{z_2}\int_x^y \omega := ⨰_{\mathbb{P}_1(\mathbb{Q}_p)} \left(\frac{t-z_2}{t-z_1}\right) d\mu_f\{x \to y\}(t) \in \mathbb{C}_p^\times. \tag{9.25}$$

Here $⨰$ denotes the multiplicative integral, in which limits of products replace the usual limits of Riemann sums, as in equation (5.8) of chapter 5.

Properties analogous to those of lemma 9.8, with addition replaced by multiplication, hold for the multiplicative integral:

LEMMA 9.9. *The double multiplicative integral of definition 9.7 satisfies the following properties:*

$$\times\!\!\!\!\!\!\int_{z_1}^{z_3}\int_x^y \omega = \times\!\!\!\!\!\!\int_{z_1}^{z_2}\int_x^y \omega \times \times\!\!\!\!\!\!\int_{z_2}^{z_3}\int_x^y \omega, \quad \text{for all } z_1, z_2, z_3 \in \mathcal{H}_p; \tag{9.26}$$

$$\times\!\!\!\!\!\!\int_{z_1}^{z_2}\int_{x_1}^{x_3} \omega = \times\!\!\!\!\!\!\int_{z_1}^{z_2}\int_{x_1}^{x_2} \omega \times \times\!\!\!\!\!\!\int_{z_1}^{z_2}\int_{x_2}^{x_3} \omega, \quad \text{for all } x_1, x_2, x_3 \in \mathbb{P}_1(\mathbb{Q}); \tag{9.27}$$

$$\times\!\!\!\!\!\!\int_{\gamma z_1}^{\gamma z_2}\int_{\gamma x}^{\gamma y} \omega = \left(\times\!\!\!\!\!\!\int_{z_1}^{z_2}\int_x^y \omega\right)^{w^{|\gamma|_p} w_\infty^{|\gamma|_\infty}}, \quad \text{for all } \gamma \in R^\times. \tag{9.28}$$

PROOF. The proof is identical to that of lemma 9.8; see exercise 4. □

9.4. Some p-adic cocycles

Given any $\tau \in \mathcal{H}_p$, guided by equation (8.5) of the previous chapter in the case $n = 1$, we define a 2-cochain $\kappa_\tau \in Z^2(\Gamma, \mathbb{C}_p^\times)$ by choosing a base point $x \in \mathbb{P}_1(\mathbb{Q})$ and setting

$$\kappa_\tau(\gamma_1, \gamma_2) = \times\!\!\!\!\!\!\int_\tau^{\gamma_1\tau}\int_{\gamma_1 x}^{\gamma_1\gamma_2 x} \omega.$$

One verifies (by the same calculations that were performed in the previous chapter) that κ_τ is a two-cocycle and that its image in $H^2(\Gamma, \mathbb{C}_p^\times)$ depends only on the form f, not on the base points x and τ that were used to define it. The following is a p-adic analogue of conjecture 8.6 of chapter 8.

CONJECTURE 9.10. *Let $q \in \mathbb{Q}_p^\times$ be Tate's p-adic period attached to E. The natural image of κ_τ in $H^2(\Gamma, \mathbb{C}_p^\times/q^{\mathbb{Z}})$ is* 0.

A part of conjecture 9.10 is established in [**Da01**], where it is shown to be related to conjectures of Mazur, Tate and Teitelbaum [**MTT**] on values of derivatives of p-adic L-functions, conjectures that were later established by Greenberg and Stevens [**GS93**]. Assume from now on that conjecture 9.10 holds for E.

COROLLARY 9.11. *There exists a one-cochain*

$$\tilde{\xi}_\tau \in C^1(\Gamma, \mathbb{C}_p^\times/q^{\mathbb{Z}})$$

such that $\kappa_\tau = d\tilde{\xi}_\tau$ (mod $q^{\mathbb{Z}}$).

We note that the the one-cochain $\tilde{\xi}_\tau$ is well-defined modulo

$$Z^1(\Gamma, \mathbb{C}_p^\times/q^{\mathbb{Z}}) = \hom(\Gamma, \mathbb{C}_p^\times/q^{\mathbb{Z}}).$$

We now invoke the following counterpart of theorem 8.9 of chapter 8:

THEOREM 9.12. *The abelianisation of Γ is finite. In particular the cohomology group $H^1(\Gamma, \mathbb{C}_p^\times/q^{\mathbb{Z}})$ is finite.*

PROOF. See [**Ih68**], chapter 3, §1-7, or exercise 9. □

Let e_Γ denote the exponent of $H^1(\Gamma, \mathbb{C}_p^\times/q^{\mathbb{Z}})$. Then the cochain

$$\xi_\tau := e_\Gamma \tilde{\xi}_\tau$$

is a well-defined element of $C^1(\Gamma, \mathbb{C}_p^\times/q^{\mathbb{Z}})$.

The cochain ξ_τ depends, of course, on the choice of base points x and τ. To analyse its dependence on x, observe that if y is another base point in $\mathbb{P}_1(\mathbb{Q})$, and κ_τ^x and κ_τ^y are the respective associated 2-cocyles, then as in the proof of lemma 8.4,

$$\kappa_\tau^x - \kappa_\tau^y = d\rho_\tau^{x,y},$$

where $\rho_\tau^{x,y} \in C^1(\Gamma, \mathbb{C}_p^\times/q^{\mathbb{Z}})$ is defined by

$$\rho_\tau^{x,y}(\gamma) = \times\!\!\!\!\!\!\int_\tau^{\gamma\tau}\int_{\gamma x}^{\gamma y}\omega.$$

In particular, this cochain vanishes on the stabiliser subgroup $\Gamma_\tau \subset \Gamma$ of τ.

It follows that the restriction θ_τ of ξ_τ to Γ_τ is independent of the choices of x and $\tilde{\xi}_\tau$ that were made in defining it.

The definition of θ_τ makes it clear that this one-cochain is in fact a well-defined homomorphism from Γ_τ to $\mathbb{C}_p^\times/q^{\mathbb{Z}}$. Hence we have associated (as in the case $n = 1$ of chapter 8) a well-defined element

$$\theta_\tau \in \hom(\Gamma_\tau, \mathbb{C}_p^\times/q^{\mathbb{Z}})$$

to any $\tau \in \mathcal{H}_p$.

9.5. Stark-Heegner points

Of course, the invariant θ_τ can only be interesting if the stabiliser Γ_τ is non-trivial, i.e., if τ belongs to $\mathcal{H}_p'$. This occurs precisely when τ belongs to $K \cap \mathcal{H}_p$, where $K \subset \mathbb{C}_p$ is a real quadratic field in which p is inert or ramified.

More specifically, we will assume that K is a real quadratic subfield of $\mathbb{C}_p$ satisfying the following "modified Heegner hypotheses" relative to N, analogous to those imposed in chapter 3, except that the roles of the places p and ∞ have been interchanged:

(1) p is inert in K;
(2) all the primes dividing M are split in K.

Note that one then has

$$S_{E,K} = \{p, \infty_1, \infty_2\} \cup \{\lambda | M\},$$

which has odd cardinality, so that $\operatorname{sign}(E, K) = -1$ in this case.

For any $\tau \in \mathcal{H}_p \cap K$ we define as in the previous section the *associated order* $\mathcal{O}_\tau$ to be the ring of matrices in $M_0(M)[1/p]$ which preserve the line spanned by the column vector $(\tau, 1)$. This order is isomorphic to a $\mathbb{Z}[1/p]$-order in K and hence the group $\mathcal{O}_{\tau,1}^\times$ of elements of $\mathcal{O}_\tau$ of determinant one has rank one. In fact, the stabiliser Γ_τ of τ in Γ is identified with $\mathcal{O}_{\tau,1}^\times/\langle\pm 1\rangle$ and is therefore a cyclic group of infinite order. Let γ_τ denote a generator of Γ_τ.

The element

$$J_\tau = \theta_\tau(\gamma_\tau) = \xi_\tau(\gamma_\tau) \in \mathbb{C}_p^\times/q^{\mathbb{Z}}$$

is a canonical invariant in $\mathbb{C}_p^\times/q^{\mathbb{Z}}$ attached to $\tau \in \mathcal{H}_p'$, a p-adic analogue of the invariant that was also denoted by J_τ in chapter 8. (Of course, the most direct analogy is with the case $n = 1$ of the construction of chapter 8.)

Let

$$P_\tau = \Phi_{\mathrm{Tate}}(J_\tau) \in E(\mathbb{C}_p),$$

and define:

$$\Phi_N'^{(p)}(\tau) := P_\tau.$$

Fixing $\tau \in \mathcal{H}_p'$, let H^+ denote the narrow ring class field of K attached to the order $\mathcal{O}_\tau$, and let H denote the usual ring class field attached to that order. The field H^+ is an extension of H of degree at most 2 which is trivial precisely when $\mathcal{O}_\tau^\times$ contains an element of norm -1. The Galois group of H^+/H is generated by complex conjugation c, defined using any complex embedding of H^+. View H and H^+ as subfields of $\mathbb{C}_p$ by fixing an embedding $H^+ \subset \mathbb{C}_p$.

CONJECTURE 9.13. *Let $\tau \in \mathcal{H}_p'$ be a special point, Then*

$$P_\tau = \Phi'^{(p)}(\tau) \text{ belongs to } E(H^+),$$

and

$$cP_\tau = w_\infty P_\tau.$$

This conjecture, like the main conjecture of chapter 8, can be made more precise by formulating a Shimura reciprocity law for the points P_τ. The reader is referred to section 5.2 of [**Da01**] where this is spelled out in detail, or better yet, invited to work through exercise 6.

9.6. Computing Stark-Heegner points

As in the constructions that were described in chapter 8, the main difficulty in making the definition of J_τ explicit and computable is the need to produce a 1-cochain $\tilde{\xi}_\tau$ satisfying the relation

$$d\tilde{\xi}_\tau = \kappa_\tau.$$

We now outline a method for computing $\tilde{\xi}_\tau$, at least in an important special case.

Recalling the notations of section 2.7, for any abelian group A with trivial Γ-action let $\mathcal{M}(A)$ denote the group of A-valued modular symbols as in definition 2.15 of chapter 2. The group Γ acts on $\mathcal{M}(A)$ by the rule

$$\gamma \cdot m\{x \rightarrow y\} = m\{\gamma^{-1}x \rightarrow \gamma^{-1}y\}.$$

Note that in general Γ does not act transitively on $\mathbb{P}_1(\mathbb{Q})$. To remedy this one chooses a base point $x \in \mathbb{P}_1(\mathbb{Q})$ and writes $\mathcal{M}_0(A)$ for the group of functions on $(\Gamma x) \times (\Gamma x)$ arising from the restriction of a modular symbol on $\mathbb{P}_1(\mathbb{Q}) \times \mathbb{P}_1(\mathbb{Q})$.

A point $\tau \in \mathcal{H}_p$ gives rise to a one-cochain $c_\tau \in C^1(\Gamma, \mathcal{M}_0(\mathbb{C}_p^\times))$ by the rule

$$c_\tau(\gamma)\{x \rightarrow y\} = \times\!\!\!\!\!\!\int_\tau^{\gamma\tau} \int_x^y \omega.$$

It can be checked by a direct computation that c_τ is a one-cocycle.

Let $\bar{c}_\tau$ be the natural image of c_τ in $H^1(\Gamma, \mathcal{M}_0(\mathbb{C}_p^\times))$. Let $\mathcal{F}(\mathbb{C}_p^\times)$ denote the space of $\mathbb{C}_p^\times$-valued functions on Γx, equipped with the natural Γ-action. Consider the exact sequence of Γ-modules

$$0 \longrightarrow \mathbb{C}_p^\times \xrightarrow{i} \mathcal{F}(\mathbb{C}_p^\times) \xrightarrow{\Delta} \mathcal{M}_0(\mathbb{C}_p^\times) \longrightarrow 0, \tag{9.29}$$

where i is the obvious inclusion into the space of constant functions, and Δ is defined by

$$\Delta g\{x \rightarrow y\} := g(y) - g(x).$$

Taking the Γ-cohomology of (9.29) yields a connecting homomorphism

$$\delta : H^1(\Gamma, \mathcal{M}_0(\mathbb{C}_p^\times)) \longrightarrow H^2(\Gamma, \mathbb{C}_p^\times),$$

and it can be checked that

$$\delta\bar{c}_\tau = \bar{\kappa}_\tau^\#,$$

where $\bar{\kappa}_\tau^\#$ is the cohomology class represented by the 2-cocycle

$$\kappa_\tau^\#(g_0, g_1) := \kappa_\tau(g_1^{-1}, g_0^{-1}).$$

It follows that $\delta\bar{c}_\tau$ is a coboundary precisely when $\bar{\kappa}_\tau$ is, and the same holds after replacing $\mathbb{C}_p^\times$ by $\mathbb{C}_p^\times/q^{\mathbb{Z}}$ for any $q \in \mathbb{C}_p^\times$. Therefore the following conjecture (made under hypothesis 9.4) is a natural strengthening of conjecture 9.10.

Conjecture 9.14. *The natural image of c_τ in $H^1(\Gamma, \mathcal{M}_0(\mathbb{C}_p^\times/q^{\mathbb{Z}}))$ is trivial.*

This conjecture implies the existence of a modular symbol

$$\tilde{\eta}_\tau \in \mathcal{M}_0(\mathbb{C}_p^\times/q^{\mathbb{Z}})$$

with the property that

$$\times\!\!\!\!\!\!\int_\tau^{\gamma\tau}\int_x^y \omega = \tilde{\eta}_\tau\{\gamma^{-1}x \to \gamma^{-1}y\} \div \tilde{\eta}_\tau\{x \to y\} \pmod{q^{\mathbb{Z}}}. \tag{9.30}$$

Of course the modular symbol $\tilde{\eta}_\tau$ is only well-defined modulo $H^0(\Gamma, \mathcal{M}_0(\mathbb{C}_p^\times/q^{\mathbb{Z}}))$. It can be shown (cf. exercise 7) that this group injects into $\hom(\Gamma, \mathbb{C}_p^\times/q^{\mathbb{Z}})$, hence is annihilated by the integer e_Γ introduced in the previous section. Setting

$$\eta_\tau = e_\Gamma \cdot \tilde{\eta}_\tau,$$

we see that this modular symbol is independent of the choice of $\tilde{\eta}_\tau$ satisfying (9.30).

We define the "semi-indefinite integral"

$$\times\!\!\!\!\!\!\int^\tau \int_x^y e_\Gamma \omega := \eta_\tau\{x \to y\} \in \mathbb{C}_p^\times/q^{\mathbb{Z}},$$

for any $\tau \in \mathcal{H}_p$ and $x, y \in \Gamma x_0$. This expression satisfies all the properties suggested by the notation (cf. exercise 8)

$$\times\!\!\!\!\!\!\int^\tau \int_x^y e_\Gamma\omega \times \times\!\!\!\!\!\!\int^\tau \int_y^z e_\Gamma\omega = \times\!\!\!\!\!\!\int^\tau \int_x^z e_\Gamma\omega, \quad \text{for all } x, y, z \in \mathbb{P}_1(\mathbb{Q}); \tag{9.31}$$

$$\times\!\!\!\!\!\!\int^{\tau_2} \int_x^y e_\Gamma\omega \div \times\!\!\!\!\!\!\int^{\tau_1} \int_x^y e_\Gamma\omega = \left(\times\!\!\!\!\!\!\int_{\tau_1}^{\tau_2} \int_x^y \omega\right)^{e_\Gamma}, \quad \text{for all } \tau_1, \tau_2 \in \mathcal{H}_p; \tag{9.32}$$

$$\times\!\!\!\!\!\!\int^{\gamma\tau} \int_{\gamma x}^{\gamma y} e_\Gamma\omega = \times\!\!\!\!\!\!\int^\tau \int_x^y e_\Gamma\omega \text{ for all } \gamma \in \Gamma. \tag{9.33}$$

More importantly, the cochain ξ_τ of the previous section can be expressed in terms of the semi-indefinite integral attached to η_τ:

Proposition 9.15. *For all $\gamma \in \Gamma$,*

$$\xi_\tau(\gamma) = \times\!\!\!\!\!\!\int^\tau \int_x^{\gamma x} e_\Gamma\omega.$$

Proof. Let ξ_τ^o denote the expression on the right hand side. In light of the property characterising ξ_τ it suffices to show that

$$d\xi_\tau^o(g_0, g_1) = \kappa_\tau(g_0, g_1)^{e_\Gamma}.$$

But

$$\begin{aligned}
d\xi_\tau^o(g_0,g_1) &= \times\!\!\!\!\!\!\int^\tau \int_x^{g_1 x} e_\Gamma \omega \div \times\!\!\!\!\!\!\int^\tau \int_x^{g_0 g_1 x} e_\Gamma \omega \times \times\!\!\!\!\!\!\int^\tau \int_x^{g_0 x} e_\Gamma \omega \\
&= \times\!\!\!\!\!\!\int^\tau \int_x^{g_1 x} e_\Gamma \omega \times \times\!\!\!\!\!\!\int^\tau \int_{g_0 g_1 x}^{g_0 x} e_\Gamma \omega \\
&= \times\!\!\!\!\!\!\int^\tau \int_x^{g_1 x} e_\Gamma \omega \div \times\!\!\!\!\!\!\int^{g_0^{-1}\tau} \int_x^{g_1 x} e_\Gamma \omega \\
&= \left(\times\!\!\!\!\!\!\int_{g_0^{-1}\tau}^{\tau} \int_x^{g_1 x} \omega \right)^{e_\Gamma} = \left(\times\!\!\!\!\!\!\int_{\tau}^{g_0\tau} \int_{g_0 x}^{g_0 g_1 x} \omega \right)^{e_\Gamma}.
\end{aligned}$$

The proposition follows. ☐

The calculation of $J_\tau = \xi_\tau(\gamma_\tau)$ is thus reduced to computing the indefinite integral

$$\times\!\!\!\!\!\!\int^\tau \int_x^{y} e_\Gamma \omega. \tag{9.34}$$

We present a method for doing this, at least in the special case where $M = 1$ so that $\Gamma = \mathbf{PSL}_2(\mathbb{Z}[1/p])$. (This corresponds to the case where E has prime conductor p.) Note that in this case one may choose $e_\Gamma = 1$ (cf. exercise 10).

First observation. By the remark in the paragraph titled "first observation" in section 2.7, it is enough to compute (9.34) when x and y are adjacent cusps in the sense of that paragraph.

Second observation: Any pair of adjacent cusps is Γ-equivalent to $(0,\infty)$. By the Γ-equivariance property of (9.33), it is enough to be able to compute the expression

$$\times\!\!\!\!\!\!\int^\tau \int_0^{\infty} e_\Gamma \omega \quad \text{(as a function of } \tau \in \mathcal{H}_p).$$

Third observation: An elementary computation (cf. [**DG01**], or exercise 11) shows that

$$\times\!\!\!\!\!\!\int^\tau \int_0^{\infty} e_\Gamma \omega = \left(\times\!\!\!\!\!\!\int_{1-\frac{1}{\tau}}^{\tau-1} \int_0^{\infty} \omega \right)^{e_\Gamma}. \tag{9.35}$$

In this way the calculation of ξ_τ is reduced to that of the double multiplicative integrals defined in section 9.3.

In more general situations (both in the S-arithmetic case of this chapter, and in the Hilbert modular setting of the previous chapter), the calculation of the cochain ξ_τ presents interesting difficulties, and it would be useful to develop feasible algorithms for performing this computation.

Further results

The theory of integration on $\mathcal{H}_p \times \mathcal{H}$ and its application to defining "Stark-Heegner points" attached to real quadratic fields is explained in [**Da01**]. Numerical evidence for conjecture 9.13, resting on the explicit definition of the map $\Phi_N'^{(p)}$

described in section 9.6, is given in [**DG01**]. Some of the ideas explained in this chapter are also covered (with a different emphasis focussing more strongly on p-adic L-functions) in [**BD01**] and [**BDG03**].

For various perspectives on the action of subgroups of $\mathbf{PSL}_2(\mathbb{Z}[1/p])$ on $\mathcal{H}_p \times \mathcal{H}$, see [**Ih68**], [**Ih79**], [**Se80**], and [**Sta87**].

Exercises

(1) Let Γ be the finite index subgroup of $\mathbf{SL}_2(\mathbb{Z}[1/p])$ consisting of the matrices which are upper triangular modulo M.
   (a) Show that this group acts on $\mathcal{H}$ and on the p-adic upper half plane $\mathcal{H}_p$ with dense orbits, but that the action of Γ on $\mathcal{H}_p \times \mathcal{H}$ is discrete.
   (b) Let $\mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1$ denote the Bruhat-Tits tree of $\mathcal{H}_p$, equipped with its natural Γ-action. Show that there are precisely two (resp one) orbits for the action of Γ on $\mathcal{T}_0$ (resp. $\mathcal{T}_1$).
   (c) Show that the stabiliser of a vertex of $\mathcal{T}$ in Γ is conjugate to the group $\Gamma_0(M)$, while the stabiliser of an edge is conjugate to the group $\Gamma_0(N)$.

(2) Let $\Gamma = \mathbf{SL}_2(\mathbb{Z}[1/p])$ be the group acting on $\mathcal{H} \times \mathcal{H}_p$ by Möbius transformations. A point $\tau \in \mathcal{H}$ is called a *special point* if its stabiliser in Γ is infinite.
   (a) Show that if τ is a special point, then τ belongs to $\mathcal{H} \cap K$, where K is a quadratic imaginary subfield of $\mathbb{C}$ in which p is split. Write $p = \mathfrak{p}\bar{\mathfrak{p}}$.
   (b) Let j be the usual modular j-function. Show that if τ is special, then $j(\tau)$ belongs to the ring $\mathcal{O}_{\bar{\mathbb{Q}}}$ of algebraic integers. (Hint: use the theory developed in chapter 3, particularly exercise (2) of that chapter.)
   (c) Let $\bar{\mathfrak{p}}$ be an ideal of $\mathcal{O}_{\bar{\mathbb{Q}}}$ above p and let

$$\mathrm{red} : \mathcal{O}_{\bar{\mathbb{Q}}} \longrightarrow \bar{\mathbb{F}}_p$$

   denote the corresponding reduction map. Given a special point τ in $\mathcal{H}$, define the weight of τ by letting α_1 and $\alpha_2 \in K$ denote the eigenvalues of a generator for the stabiliser subgroup Γ_τ, and setting

$$\mathrm{wt}(\tau) = \max(|\mathrm{ord}_{\mathfrak{p}}(\alpha_1)|, |\mathrm{ord}_{\mathfrak{p}}(\alpha_2)|).$$

   Letting $n = \mathrm{wt}(\tau)$, show that

$$\eta(\tau) := \mathrm{red}(j(\tau))$$

   belongs to $\mathbb{F}_{p^n}$ and to no smaller extension of $\mathbb{F}_p$, where n is the weight of τ.
   (d) * Let $SS \subset \mathbb{P}_1(\mathbb{F}_{p^2})$ denote the set of supersingular j-invariants in characteristic p, and let $\mathcal{H}'$ denote the set of special points of $\mathcal{H}$. Show that the map

$$\eta : \mathcal{H}'/\Gamma \longrightarrow \bar{\mathbb{F}}_p - SS$$

   is a bijection. (See also [**Ih68**], chapter 5.)

(3) Prove lemma 9.2 in the text.

(4) Prove that the double multiplicative integral satisfies the properties given in lemma 9.9.

(5) Let $\mathcal{H}_p^{\mathrm{unram}}$ denote the set of points in $\mathcal{H}_p$ which map to $\mathcal{T}_0$ under the reduction map defined in chapter 5. Define a $\mathbb{C}$-valued double integral attached to a form

in $S_2(\mathcal{T},\Gamma)$ by the rule

$$\times\!\!\!\!\int_{\tau_1}^{\tau_2}\int_{\tau_3}^{\tau_4}\omega := \sum_{e:r(\tau_3)\to r(\tau_4)}\int_{\tau_1}^{\tau_2} f_e(z)dz, \quad \tau_1,\tau_2\in\mathcal{H}, \quad \tau_3,\tau_4\in\mathcal{H}_p^{\mathrm{unram}},$$

the the sum being taken over the ordered edges in the path joining $r(\tau_3)$ to $r(\tau_4)$.

(a) Show that this integral satisfies all the formal properties of the $\mathbb{C}_p^\times$-valued integral stated in lemma 9.8.

(b) Mimic the constructions of chapters 8 and 9 to define a map

$$\Phi'_N : \mathcal{H}'/\Gamma \longrightarrow E(\mathbb{C}).$$

(Hint: After defining κ_τ in the obvious way, the analogue of conjectures 8.6 and 9.10 can be proved using theorem 2.19 of section 2.7. This should yield an *explicit formula* for a one-cochain $\xi_\tau \in \mathrm{C}^1(\Gamma,\mathbb{C}/\Lambda_E)$ satisfying $d\xi_\tau = e_\Gamma\kappa_\tau \pmod{\Lambda_E}$).)

(c) State the analogue of conjecture 9.13 in this setting, and prove it using the theory developed in chapter 3.

(For more details on this exercise see section 4 of [**BDG03**].)

(6) Formulate precisely a Shimura reciprocity law in the style of the conjectures of chapter 8 for the points P_τ defined in this chapter.

(7) Construct an injective homomorphism from $H^0(\Gamma,\mathcal{M}_0(A))$ to $\hom(\Gamma,A)$ (for any abelian group A).

(8) Show that the indefinite multiplicative integral satisfies the properties listed in equations (9.31), (9.32), and (9.33). (Conclude that the same is true for the properties listed in lemma 9.8.)

(9) Let $\Gamma \subset \mathbf{PSL}_2(\mathbb{Z}[1/p])$ be the group of matrices which are upper triangular modulo M. Let A be any Γ-module, and let $\mathcal{F}(\mathcal{T}_0,A)$ and $\mathcal{F}(\mathcal{T}_1,A)$ denote the module of A-valued functions on the set of vertices and unordered edges of $\mathcal{T}$ respectively.

(a) Construct an exact sequence of Γ-modules

$$0 \longrightarrow A \longrightarrow \mathcal{F}(\mathcal{T}_0,A) \longrightarrow \mathcal{F}(\mathcal{T}_1,A) \longrightarrow 0.$$

(b) Let $\Gamma_0'(M) = \alpha_p\Gamma_0(M)\alpha_p^{-1}$, where α_p is the element of Γ of determinant p satisfying $\alpha_p e_o = \bar{e}_o$. Construct an exact sequence

$$\cdots \longrightarrow H^i(\Gamma_0(M),A)\oplus H^i(\Gamma_0'(M),A) \longrightarrow H^i(\Gamma_0(N),A) \xrightarrow{\delta}$$
$$\longrightarrow H^{i+1}(\Gamma,A) \longrightarrow H^{i+1}(\Gamma_0'(M),A)\oplus H^{i+1}(\Gamma_0'(M),A) \longrightarrow \cdots$$

(Hint: Take the Γ-cohomology of the exact sequence displayed in (a) and apply Shapiro's lemma.)

(c) Use this exact sequence to analyse the abelianisation of Γ.

(10) Let $\Gamma = \mathbf{PSL}_2(\mathbb{Z}[1/p])$. Show that the space $\mathcal{M}(A)^\Gamma$ is trivial, for any abelian group A. Compute the abelianisation of Γ in this case.

(11) Prove equation (9.35) in the text.

(12) Let $\mathcal{O} = \mathbb{Z}[\tau]$ be the real quadratic order defined by

$$\tau = \begin{cases} \sqrt{n^2+1} & \text{if } n \geq 2 \text{ is even;} \\ \frac{n+\sqrt{n^2+4}}{2} & \text{if } n \text{ is odd.} \end{cases}$$

This order has fundamental unit η of norm -1 given by

$$\eta = \begin{cases} n + \sqrt{n^2+1} & \text{if } n \geq 2 \text{ is even;} \\ \frac{n+\sqrt{n^2+4}}{2} & \text{if } n \text{ is odd.} \end{cases}$$

Show that

$$J_\tau = \times\!\!\!\!\!\!\int_{-\eta}^{\eta'}\int_0^\infty \omega,$$

where η' is the Galois conjugate of η.

CHAPTER 10

Kolyvagin's theorem

Let E be an elliptic curve over $\mathbb{Q}$ and let K be a quadratic extension of $\mathbb{Q}$ satisfying $\operatorname{sign}(E,K) = -1$. A theme of the last few chapters has been the construction of so-called "Heegner Systems" attached to the pair (E,K). Recall that if $\{P_n\}_{(n,N)=1}$ is such a Heegner system, one sets $P_K := \operatorname{Trace}_{H/K}(P_1)$.

The goal of this chapter is to explain the main ideas behind the proof of Kolyvagin's Theorem whose statement, given in chapter 3, we now recall:

THEOREM 10.1 (Kolyvagin). *If P_K is a point of infinite order, then*

(1) *The rank of $E(K)$ is equal to 1.*
(2) *The Shafarevich-Tate group of E over K is finite.*

Like the proof of the weak Mordell-Weil theorem sketched in chapter 1, the proof of theorem 10.1 proceeds by studying $E(K)/pE(K)$ – or rather, the p-Selmer group $\operatorname{Sel}_p(E/K)$ – for a suitably chosen "descent prime" p. It will be convenient to assume that p satisfies the following conditions:

(1) p does not divide $6N$;
(2) The natural homomorphism from $G_{\mathbb{Q}}$ to $\operatorname{Aut}(E_p) \simeq \mathbf{GL}_2(\mathbb{F}_p)$ is surjective.

We will then prove the following "mod p" version of Kolyvagin's theorem:

THEOREM 10.2. *If the image of P_K in $E(K)/pE(K)$ is non-zero, then the p-Selmer group $\operatorname{Sel}_p(E/K)$ is generated by $\delta(P_K)$. In particular $E(K)$ has rank one and $Ш(E/K)_p$ is trivial.*

REMARK 10.3. Theorem 10.2 does not imply the full strength of theorem 10.1, because of the restrictions that are made on p. However, if E has no complex multiplications, these restrictions exclude only *finitely many* primes p, by a theorem of Serre [**Se72**]:

THEOREM 10.4 (Serre). *Let E be an elliptic curve defined over a number field F. Assume that E has no complex multiplications, i.e., that $End_{\bar{F}}(E) = \mathbb{Z}$. Then the natural homomorphism from G_F to $Aut(E_p)$ is surjective, for all but finitely many primes p.*

REMARK 10.5. Note that theorem 10.2 applied to a single prime p which does not divide P_K and satisfies conditions (1) and (2) above yields the most interesting consequence of Kolyvagin's theorem: for instance it is enough to prove part (1) of theorem 10.1. It therefore makes sense to focus on its proof, which conveys the main ideas while avoiding some technical complications that appear in the full proof of theorem 10.1. This is what will be done in the remainder of chapter 10.

10.1. Bounding Selmer groups

We begin by presenting a general approach for bounding the size of Selmer groups which plays an important role in Kolyvagin's argument (and also in Wiles' proof of the Shimura-Taniyama-Weil conjecture; for this latter point, see [**DDT95**], section 2.3).

Let K be any number field and let $G_K = \mathrm{Gal}(\bar{K}/K)$ denote its absolute Galois group endowed with the Krull topology. Let M be any finite module equipped with a continuous action of G_K. Recall the Galois cohomology group

$$H^1(K, M) = H^1(G_K, M),$$

defined as the group of continuous one-cocycles on G_K with values in M, modulo the group of 1-coboundaries on G_K.

A prime v of K is said to be *good* for M if

(1) M is unramified at v;
(2) v does not divide the cardinality of M.

Let I_v be the inertia subgroup of $G_v = G_{K_v}$, viewed as a subgroup of G_K by fixing an embedding of $\bar{K}_v$ into $\bar{K}$. Let K_v^{nr} denote the maximal unramified extension of K_v; its Galois group is identified with G_v/I_v and is isomorphic to $\hat{\mathbb{Z}}$, with a canonical topological generator given by the frobenius element Frob_v at v.

If v is good, one disposes of the short exact inflation-restriction sequence

$$0 \longrightarrow H^1(K_v^{\mathrm{nr}}/K_v, M) \xrightarrow{\mathrm{inf}} H^1(K_v, M) \xrightarrow{\partial_v} H^1(I_v, M)^{G_{K_v}} \longrightarrow 0.$$

The image of the group $H^1(K_v^{\mathrm{nr}}/K_v, M)$ under inflation is called the *finite* or *unramified* part of $H^1(K_v, M)$, and is denoted $H^1_f(K_v, M)$. The quotient

$$H^1_s(K_v, M) := H^1(K_v, M)/H^1_f(K_v, M) = H^1(I_v, M)^{G_{K_v}}$$

is called the *singular part* or *singular quotient* of $H^1(K_v, M)$. Following a terminology suggested by Mazur, the natural projection

$$\partial_v : H^1(K_v, M) \longrightarrow H^1_s(K_v, M)$$

will be referred to as the *residue map* at v. If $c \in H^1(K, M)$ is a global cohomology class, we will also denote by c_v its natural image in $H^1(K_v, M)$ under the restriction map to G_v, and set, by abuse of notation, $\partial_v(c) := \partial_v(c_v)$.

Definition 10.6. A set of *Selmer conditions* attached to M and K is a collection of subgroups $\mathcal{L}_v \subset H^1(K_v, M)$ for each place v of K, such that

$$\mathcal{L}_v = H^1_f(K_v, M) \text{ for all but finitely many } v.$$

Note that this definition makes sense, since for a given M all but finitely many places v are good and therefore the groups $H^1_f(K_v, M)$ are defined for them.

Definition 10.7. Let $\mathcal{L} = \{\mathcal{L}_v\}_v$ be a set of Selmer conditions attached to M and K. The *Selmer group* attached to the triple $(K, M, \{\mathcal{L}_v\})$, denoted $H^1_{\mathcal{L}}(K, M)$, is the set of $c \in H^1(K, M)$ such that c_v belongs to $\mathcal{L}_v$, for all v.

The principal example of interest to us is the one where $M = E_p$ is the module of p-division points of an elliptic curve and where

$$\mathcal{L}_v = \delta_v(E(K_v)/pE(K_v)).$$

The fact that $\mathcal{L}_v$ is a set of Selmer conditions follows from the proof of proposition 1.7 (more precisely, the part of that proof that is developed in exercise 7 of chapter 1). By definition,

$$H^1_{\mathcal{L}}(K, E_p) = \mathrm{Sel}_p(E/K),$$

where $\mathrm{Sel}_p(E/K)$ is the p-Selmer group of chapter 1 attached to E/K.

We have the following general finiteness property for Selmer groups:

PROPOSITION 10.8. *If $\mathcal{L}$ is any set of Selmer conditions for M, then*

$$\#H^1_{\mathcal{L}}(K,M) < \infty.$$

PROOF. This is a direct generalisation of exercise 8 and is left to the reader. $\square$

We now present a technique for bounding the orders of Selmer groups, whereby this question is turned into the problem of manfacturing classes in $H^1(K,M)$ with prescribed residues.

For this it is necessary to introduce some further notions arising from Tate local duality in Galois cohomology. Let $M^* := \hom(M, \mathbf{G}_m)$ denote the *Kummer dual* of M, equipped with its natural G_K-action

$$(\sigma f)(m) := \sigma f(\sigma^{-1}m), \quad \text{for } \sigma \in G_K, \quad f \in M^*, \quad m \in M$$

arising from the action of G_K on M and on the roots of unity. The cup product combined with the calculation of the local Brauer group in local class field theory yields a pairing

$$\langle\ ,\ \rangle_v : H^1(K_v, M) \times H^1(K_v, M^*) \longrightarrow H^2(K_v, G_m) = \mathbb{Q}/\mathbb{Z}.$$

THEOREM 10.9 (Tate). *The pairing $\langle\ ,\ \rangle_v$ is a non-degenerate bilinear pairing. If v is a good prime for M, it is also good for M^* and the groups $H^1_f(K_v, M)$ and $H^1_f(K_v, M^*)$ are orthogonal complements of each other under this pairing.*

PROOF. For a discussion and proof of this theorem, see [**Ta62**], §2 and [**Mi86**], ch, I. (See also exercise 3 for a discussion in the case where v is a good prime for M.) $\square$

Given $\mathcal{L}_v \subset H^1(K_v, M)$, let $\mathcal{L}_v^* \subset H^1(K_v, M^*)$ denote the exact annihilator of $\mathcal{L}_v$ under the local Tate pairing. It follows immediately from Tate's theorem that if the groups $\mathcal{L}_v$ form a set of Selmer conditions for M, then the groups $\mathcal{L}_v^*$ are a collection of Selmer conditions for $H^1(K, M^*)$.

DEFINITION 10.10. The Selmer group attached to M^* and $\mathcal{L}^* := \{\mathcal{L}_v^*\}_v$ is called the *dual Selmer group* of $H^1_{\mathcal{L}}(K,M)$.

While the orders of $H^1_{\mathcal{L}}(K,M)$ and its dual Selmer group $H^1_{\mathcal{L}^*}(K,M^*)$ are subtle global invariants which are typically difficult to compute, the ratio of these orders is a product of simple local terms which can be calculated in practice without difficulty. This is the content of the following key result.

THEOREM 10.11 (Duality theorem for Selmer groups).

$$\frac{\#H^1_{\mathcal{L}}(K,M)}{\#H^1_{\mathcal{L}^*}(K,M^*)} = \frac{\#H^0(K,M)}{\#H^0(K,M^*)} \prod_v \frac{\#\mathcal{L}_v}{\#H^0(K_v,M)}.$$

See [**DDT95**], §2.3 for an outline of the proof, which relies on the full force of global class field theory, and specifically on the Poitou-Tate nine-term exact sequence given in thm. 4.10 of ch. I of [**Mi86**].

Note the analogy between theorem 10.11 and the Riemann-Roch theorem. For this reason the term appearing on the right-hand side of theorem 10.11 will be denoted $\chi_{\mathcal{L}}(K,M)$ and called the *Euler characteristic* attached to $H^1_{\mathcal{L}}(K,M)$.

Fix a collection $\{\mathcal{L}_v\}$ of Selmer conditions for M. Let S be any finite set of good primes for M, chosen so that

$$\mathcal{L}_v = H^1_f(K_v, M) \quad \text{for all } v \in S.$$

DEFINITION 10.12. The *relaxed* Selmer group at S, denoted $H^1_{(S)}(K, M)$, is the set of classes $c \in H^1(K, M)$ such that

$$c_v \text{ belongs to } \mathcal{L}_v, \quad \text{for all } v \notin S.$$

DEFINITION 10.13. The *restricted* Selmer group at s, denoted $H^1_{[S]}(K, M)$, is the set of classes in $c \in H^1_{\mathcal{L}}(K, M)$ such that

$$c_v = 0 \quad \text{for all } v \in S.$$

There are obvious inclusions

$$H^1_{[S]}(K, M) \subset H^1_{\mathcal{L}}(K, M) \subset H^1_{(S)}(K, M).$$

It is also clear that $H^1_{(S)}(K, M)$ and $H^1_{[S]}(K, M^*)$ are dual Selmer groups in the sense of definition 10.10. Hence theorem 10.11, applied to both $H^1_{\mathcal{L}}(K, M)$ and $H^1_{(S)}(K, M)$ yields the following useful identity:

$$\frac{\#H^1_{(S)}(K, M)}{\#H^1_{[S]}(K, M^*)} = \chi_{\mathcal{L}}(K, M) \cdot \# \left(\oplus_{v \in S} H^1_s(K_v, M)\right). \tag{10.1}$$

A finite set S of good primes for M^* is said to *control* the Selmer group $H^1_{\mathcal{L}^*}(K, M^*)$ if $H^1_{[S]}(K, M^*) = 0$, i.e., if the natural map

$$H^1_{\mathcal{L}^*}(K, M^*) \longrightarrow \oplus_{v \in S} H^1_f(K_v, M^*)$$

is injective.

The following proposition is a key ingredient in the proof of Kolyvagin's theorem, as well as in many of the arguments which bound orders of Selmer groups:

THEOREM 10.14. *Suppose that $\chi_{\mathcal{L}}(K, M) = 1$, and let S be a set of good primes which controls $H^1_{\mathcal{L}^*}(K, M^*)$. Then the cardinality of $H^1_{\mathcal{L}}(K, M)$ is equal to the cardinality of the cokernel of the residue map $\partial_S := \oplus_{v \in S} \partial_v$:*

$$\partial_S : H^1_{(S)}(K, M) \longrightarrow \oplus_{v \in S} H^1_s(K_v, M).$$

PROOF. Consider the tautological exact sequence

$$0 \longrightarrow H^1_{\mathcal{L}}(K, M) \longrightarrow H^1_{(S)}(K, M) \xrightarrow{\partial_S} \oplus_{v \in S} H^1_s(K_v, M). \tag{10.2}$$

The assumptions that $\chi_{\mathcal{L}}(K, M) = 1$ and that S controls $H^1_{\mathcal{L}^*}(K, M^*)$ implies, in light of (10.1), that the two groups appearing on the right of (10.2) have the same cardinality; the result follows. □

Thanks to theorem 10.14, the problem of bounding $H^1_{\mathcal{L}}(K, M)$ is reduced to that of producing, for a well-chosen set S of primes which controls $H^1_{\mathcal{L}^*}(K, M^*)$, a supply of cohomology classes in $H^1_{(S)}(K, M)$ whose residues at the primes of S can be controlled so that the order of the cokernel of ∂_S can be estimated. In Euler system arguments the goal is often to estimate the cokernel of ∂_S in terms of a quantity connected with the behaviour of a suitable L-function attached to M and K. (See for instance [**Da02**] for a discussion of this point of view.)

10.2. Kolyvagin cohomology classes

Returning to the case which is germane to the proof of Kolyvagin's theorem, let E be an elliptic curve over any number field K, and let p be a prime of good reduction for E. One sets

$$M = E_p, \quad \mathcal{L}_v = \delta_v(E(K_v)/pE(K_v)),$$

so that by definition $H^1_{\mathcal{L}}(K, M) = \mathrm{Sel}_p(E/K)$, the p-Selmer group attached to E that was defined in chapter 1. Write $\mathrm{Sel}^*_p(E/K) := H^1_{\mathcal{L}^*}(K, E_p)$ for the dual Selmer group attached to $\mathrm{Sel}_p(E/K)$. In this case, the non-degenerate alternating Weil pairing $E_p \times E_p \longrightarrow \mu_p$ yields an *identification* of E_p with its Kummer dual E^*_p, so that both $\mathrm{Sel}_p(E/K)$ and $\mathrm{Sel}^*_p(E/K)$ can be viewed as contained in $H^1(K, E_p)$.

PROPOSITION 10.15. *The Euler characteristic $\chi_{\mathcal{L}}(K, E_p)$ is equal to 1. (Hence the groups $\mathrm{Sel}_p(E/K)$ and $\mathrm{Sel}^*_p(E/K)$ have the same cardinality.)*

PROOF. The global term $\#H^0(K, E_p)/\#H^0(K, E^*_p)$ appearing in the formula for $\chi_{\mathcal{L}}(K, E_p)$ of theorem 10.11 is equal to 1 since E_p and E^*_p are isomorphic as G_K-modules. (In fact, by the assumptions that were made on p, both the numerator and the denominator in this expression are equal to 1.) Note that the local term $\chi_v := \#\mathcal{L}_v/\#H^0(K_v, E_p)$ is of the form

$$\#(G \otimes \mathbb{F}_p)/\#G[p], \quad \text{where } G = E(K_v). \tag{10.3}$$

The function which to an abelian group G associates the expression in (10.3) is additive in exact sequences and trivial on finite groups. If v does not divide $p\infty$, then $E(K_v)$ is an extension of a finite group by a pro-ℓ group, with $\ell \neq p$. Hence

$$\chi_v = 1, \quad \text{for all } v \not| p\infty. \tag{10.4}$$

Let d be the degree of K over $\mathbb{Q}$. By a direct calculation,

$$\prod_{v|\infty} \chi_v = p^{-d}. \tag{10.5}$$

Finally, if v divides p, then $\prod_{v|p} E(K_v)$ is isomorphic (by the theory of the formal group logarithm; cf. [**Si86**], ch. IV and VII) to an extension of a finite group by a group which is abstractly isomorphic to $\mathbb{Z}_p^d$. Hence

$$\prod_{v|p} \chi_v = p^d. \tag{10.6}$$

Equations (10.4), (10.5) and (10.6) imply that $\chi_{\mathcal{L}}(K, E_p) = 1$, as was to be shown. $\square$

REMARK 10.16. The Selmer group $\mathrm{Sel}_p(E/K)$ and its dual $\mathrm{Sel}^*_p(E/K)$, viewed as subgroups of $H^1(K, E_p)$, are in fact *equal*. The most natural way to prove this is by local methods, by showing that $\mathcal{L}_v$ is its own orthogonal complement under the local Tate pairing on $H^1(K_v, E_p)$. (When v divides neither p nor the conductor of E, so that v is a good prime for E_p, this fact has already been explained in a more general setting.) We will not insist on the equality of $\mathrm{Sel}_p(E/K)$ and its dual, since we will not make use of it in the proof of theorem 10.2. In fact, maintaining the notational distinction between $\mathrm{Sel}_p(E/K)$ and $\mathrm{Sel}^*_p(E/K)$ is more illustrative of how a general Euler system argument for bounding $H^1_{\mathcal{L}}(K, M)$ might proceed, in cases where M is not identified with its Kummer dual. (Such as the study of the adjoint or symmetric square representations attached to E_p that arises in Wiles' work on the Shimura-Taniyama-Weil conjecture.)

For the remainder of this chapter, let us specialise further to the case where E is an elliptic curve over $\mathbb{Q}$ and K is a quadratic field satisfying $\mathrm{sign}(E,K) = -1$. Let p be a prime satisfying the conditions preceeding the statement of theorem 10.2.

DEFINITION 10.17. A rational prime ℓ is called a *Kolyvagin prime* relative to (E, K, p) if

(1) ℓ does not divide $2N\mathrm{Disc}(K)p$, and is inert in K;
(2) p divides the fourier coefficient a_ℓ exactly;
(3) p^2 divides $\ell + 1$;
(4) if K is a real quadratic field, then the fundamental unit u_K of $\mathcal{O}_K^\times$ is a p^2-th power in $(\mathcal{O}_K/\ell)^\times$.

The following proposition asserts that there are plenty of Kolyvagin primes. In fact, sufficiently many can be produced to control $\mathrm{Sel}_p^*(E/K)$.

PROPOSITION 10.18. *There are infinitely many Kolyvagin primes. In fact, there exists a finite set S of Kolyvagin primes which controls* $\mathrm{Sel}_p^*(E/K)$.

SKETCH OF PROOF. Let $M = K(E_{p^2}, u_k^{1/p^2})$ be the finite extension of $\mathbb{Q}$ obtained by adjoining to K the coordinates of the points of order p^2 in E and the p^2-th roots of u_K if K is real quadratic. The condition that ℓ be a Kolyvagin prime can be rephrased as a condition on the Frobenius element at ℓ in the Galois extension $M/\mathbb{Q}$ (cf. exercise 4) and so it follows that there are infinitely many Kolyvagin primes: in fact, these form a subset of the primes of $\mathbb{Q}$ having positive Dirichlet density. The construction of a finite set S of Kolyvagin primes which controls $\mathrm{Sel}_p^*(E/K)$ also follows from a careful application of the Chebotarev density theorem, this time to the finite extension M_S of M which is "cut out" by all the elements in $\mathrm{Sel}_p^*(E/K)$. The details of the proof are involved and the reader may wish to consult [**BD03**], theorem 3.4, or better yet, work through exercise 4 where they are spelled out. $\square$

We now describe the construction of certain cohomology classes $\kappa(\ell)$ indexed by the Kolyvagin primes. Let H_ℓ denote the ring class field of K of conductor ℓ, and set

$$G_\ell = \mathrm{Gal}(H_\ell/H), \qquad \tilde{G}_\ell = \mathrm{Gal}(H_\ell/K).$$

Note that G_ℓ is a cyclic group which is canonically identified, via class field theory, with

$$G_\ell = (\mathcal{O}_K/\ell)^\times/(\mathbb{Z}/\ell\mathbb{Z})^\times\langle u_K\rangle.$$

It follows from the definition of a Kolyvagin prime that $n_\ell = \#G_\ell$ is divisible by p^2. Let $\tilde{N}_1$ be an arbitrary lift to $\mathbb{Z}[\tilde{G}_\ell]$ of the norm element $\sum_{\sigma\in\mathrm{Gal}(H/K)}\sigma$ under the natural projection $\mathbb{Z}[\tilde{G}_\ell] \longrightarrow \mathbb{Z}[\mathrm{Gal}(H/K)]$. Denote by σ_ℓ a generator of G_ℓ, and define the following elements in $\mathbb{Z}[G_\ell]$ and $\mathbb{Z}[\tilde{G}_\ell]$:

$$N_\ell = \sum_{\sigma\in G_\ell}\sigma, \quad D_\ell = \sum_{i=0}^{n_\ell - 1} i\sigma_\ell^i, \quad \tilde{N}_\ell = N_\ell\tilde{N}_1, \quad \tilde{D}_\ell = D_\ell\tilde{N}_1. \tag{10.7}$$

Note that these elements are related by the following basic identities

$$(\sigma_\ell - 1)D_\ell = n_\ell - N_\ell, \quad (\sigma_\ell - 1)\tilde{D}_\ell = n_\ell\tilde{N}_1 - \tilde{N}_\ell.$$

All four of the group ring elements defined in (10.7) act as $\mathbb{Z}$-linear operators on $E(H_\ell)$. Let $P_\ell \in E(H_\ell)$ be the point of level ℓ in the Heegner system attached to (E, K), and set

$$Q_\ell = \tilde{D}_\ell P_\ell.$$

PROPOSITION 10.19. *The natural image of Q_ℓ in $(E(H_\ell)/pE(H_\ell))$ is invariant under the Galois group $\tilde{G}_\ell$.*

PROOF. The image of the point $D_\ell P_\ell$ in $E(H_\ell)/pE(H_\ell)$ is invariant under G_ℓ. To see this, note that p divides both n_ℓ and a_ℓ by the definition of a Kolyvagin prime, and that

$$(\sigma_\ell - 1)D_\ell P_\ell = (n_\ell - N_\ell)P_\ell = n_\ell P_\ell - a_\ell P_1. \tag{10.8}$$

The image of Q_ℓ in $E(H_\ell)/pE(H_\ell)$ is the sum of the $\mathrm{Gal}(H/K)$-translates of the image of $D_\ell P_\ell$ in the group $(E(H_\ell)/pE(H_\ell))^{G_\ell}$. Hence it is invariant under $\tilde{G}_\ell$. □

REMARK 10.20. Applying the operator $\tilde{N}_1$ to (10.8) gives:

$$(\sigma_\ell - 1)Q_\ell = n_\ell \tilde{N}_1 P_\ell - a_\ell P_K. \tag{10.9}$$

Because $E_p(H_\ell)$ is trivial (cf. exercise 1), multiplication by p is injective on $E(H_\ell)$. Thus one can define a cohomology class $\bar{\kappa}(\ell) \in H^1(\tilde{G}_\ell, E(H_\ell))_p$ by the rule

$$\bar{\kappa}(\ell)(\sigma) = \frac{(\sigma - 1)Q_\ell}{p}. \tag{10.10}$$

Let $\bar{\kappa}(\ell)$ also denote, by abuse of notation, the natural image of this class in $H^1(K, E)_p$ under inflation. Finally, denote by $\kappa(\ell)$ any lift of $\bar{\kappa}(\ell)$ to $H^1(K, E_p)$ by the map arising in the Kummer descent exact sequence. Note that the residues of $\kappa(\ell)$ depend only on $\bar{\kappa}(\ell)$ – more precisely, $\partial_v(\kappa(\ell)) \neq 0$ if and only if $\bar{\kappa}(\ell)_v \neq 0$.

PROPOSITION 10.21. *The class $\kappa(\ell)$ belongs to $H^1_{(\ell)}(K, E_p)$, the relaxed Selmer group at ℓ. Furthermore*

1. *$\partial_\ell(\kappa(\ell)) \neq 0$ if and only if the image of P_K in $E(K_\ell)/pE(K_\ell)$ is non-zero.*
2. *The class $\kappa(\ell)$ belongs to $H^1(K, E_p)^\epsilon$, where $\epsilon = \mathrm{sign}(E, \mathbb{Q})$.*

PROOF. The fact that $\partial_v(\kappa(\ell)) = \bar{\kappa}(\ell)_v = 0$ when $v \nmid \ell$ follows from the fact that H_ℓ/K is unramified outside ℓ, so that the restriction of the cocycle $\bar{\kappa}(\ell)$ to any inertia group outside ℓ is identically 0.

To prove (1), note that σ_ℓ generates the decomposition group at ℓ in $\mathrm{Gal}(H_\ell/K)$. Hence $\partial_\ell(\kappa(\ell))$ is non-zero if and only if $\bar{\kappa}(\ell)(\sigma_\ell)$ is non-zero in $E(K_\ell)/pE(K_\ell)$. But by equation (10.9),

$$\bar{\kappa}(\ell)(\sigma_\ell) = \frac{(\sigma_\ell - 1)Q_\ell}{p} = -\frac{a_\ell}{p}P_K \text{ in } E(K_\ell)/pE(K_\ell),$$

the second equality holding because p^2 divides n_ℓ. Since p divides a_ℓ exactly, the result follows. To prove part (2), one uses the fact that if τ is any lift of the generator of $\mathrm{Gal}(K/\mathbb{Q})$ to $\mathrm{Gal}(H_\ell/\mathbb{Q})$, then since this latter group is a generalised dihedral group,

$$\tau\sigma = \sigma^{-1}\tau \quad \text{for all } \sigma \in \tilde{G}_\ell.$$

Hence, by exercise 5,

$$\tau D_\ell = -D_\ell \tau \pmod{n_\ell}.$$

By definition of a Heegner system (cf. definition 3.12 of chapter 3, particularly the property that is asserted in proposition 3.11)

$$\tau D_\ell P_\ell = -D_\ell \tau P_\ell = \epsilon\sigma D_\ell P_\ell \pmod{n_\ell E(H_\ell)}$$

for some $\sigma \in \tilde{G}_\ell$. Part (2) follows after applying $\tilde{N}_1$ to this identity. □

In addition to the classes $\kappa(\ell)$ depending on a single Kolyvagin prime ℓ, it is important to have at one's disposal certain classes $\kappa(\ell_1\ell_2)$ depending (in a symmetrical way) on two distinct Kolyvagin primes ℓ_1 and ℓ_2. These are defined exactly in the same way as the classes $\kappa(\ell)$ above, but using the point

$$Q_{\ell_1\ell_2} = D_{\ell_1} D_{\ell_2} \tilde{N}_1 P_{\ell_1\ell_2}$$

arising from the point $P_{\ell_1\ell_2}$ in the Heegner system. The properties of this class are summarised below:

PROPOSITION 10.22. *The class $\kappa(\ell_1\ell_2)$ belongs to $H^1_{(\ell_1\ell_2)}(K, E_p)$, the relaxed Selmer group at $\{\ell_1, \ell_2\}$. Furthermore,*

(1) *$\partial_{\ell_2}(\kappa(\ell_1\ell_2)) \neq 0$ if and only if $v_{\ell_2}(\kappa(\ell_1)) \neq 0$.*
(2) *$\kappa(\ell_1\ell_2)$ belongs to $H^1(K, E_p)^{-\epsilon}$, where $\epsilon =$sign$(E, \mathbb{Q})$.*

The proof of proposition 10.22 is identical to that of proposition 10.21 and is left as an exercise. (Cf. exercise 6.)

10.3. Proof of Kolyvagin's theorem

The following strengthening of proposition 10.18 is needed in the proof of Kolyvagin's theorem:

PROPOSITION 10.23. *There exists a finite set $S = \{\ell_1, \ldots, \ell_t)$ of Kolyvagin primes with the property that*

(1) *S controls $\mathrm{Sel}^*_p(E/K)$.*
(2) *The image of P_K in $E(K_{\ell_i})/pE(K_{\ell_i})$ is non-zero, for $j = 1, \ldots, t$.*
(3) *The value $v_{\ell_j}(\kappa(\ell_1))$ is non-zero, for $j = 2, \ldots, t$.*

PROOF. This follows from the ideas detailed in exercise 4. □

END OF PROOF OF KOLYVAGIN'S THEOREM. Let S be a set of primes satisfying the conclusion of proposition 10.23. By property (1) stated in this proposition, combined with theorem 10.14, it is enough to bound from above the cokernel of the map

$$\partial_S : H^1_{(S)}(K, E_p) \longrightarrow \oplus_{\ell \in S} H^1_s(K_\ell, E_p). \tag{10.11}$$

Note that if ℓ is a Kolyvagin prime, the Frobenius element σ_ℓ attached to ℓ in $\mathrm{Gal}(\bar{K}/K)$ acts trivially on E_p, so that $E_p(K_\ell) = E_p$. A direct calculation (cf. part (c) of exercise 3) shows that

$$H^1_f(K_\ell, E_p) = E_p/(\sigma_\ell - 1)E_p = E_p, \qquad H^1_s(K_\ell, E_p) = \hom_{G_{K_\ell}}(\mu_p, E_p) = E_p.$$

Hence the group on the right of (10.11) is an $\mathbb{F}_p$-vector space of dimension $2t$, and each eigenspace under the action of τ has dimension t over $\mathbb{F}_p$. We bound the cokernel of ∂_S one eigenspace at a time.

Step 1. Restricting the map ∂_S to the ϵ-eigencomponents,

$$\partial_S^\epsilon : H^1_{(S)}(K, E_p)^\epsilon \longrightarrow \oplus_{j=1}^t H^1_s(K_{\ell_j}, E_p)^\epsilon,$$

one sees that the vectors $\partial_S^\epsilon(\kappa_{\ell_1}), \dots, \partial_S^\epsilon(\kappa_{\ell_t})$ are linearly independent, by proposition 10.21 and property (2) of proposition 10.23. Hence ∂_S^ϵ is surjective; it follows that $\mathrm{Sel}_p(E/K)^\epsilon$ is trivial. In particular $E(\mathbb{Q})$ (resp. $E(K)/E(\mathbb{Q})$) is finite if $\mathrm{sign}(E,\mathbb{Q}) = 1$ (resp. if $\mathrm{sign}(E,\mathbb{Q}) = -1$), which is consistent with the Birch and Swinnerton-Dyer conjecture over $\mathbb{Q}$.

Step 2. By proposition 10.22 and property (3) of proposition 10.23, it likewise follows that the $t-1$ vectors $\partial_S(\kappa(\ell_1\ell_2)), \dots, \partial_S(\kappa(\ell_1\ell_t))$ are linearly independent and belong to the $-\epsilon$-eigenspace for the action of $\mathrm{Gal}(K/\mathbb{Q})$ on $\oplus_{\ell\in S} H^1_s(K_\ell, E_p)$. Hence the map $\partial_S^{-\epsilon}$ has a cokernel of dimension at most 1. It follows that $\mathrm{Sel}_p(E/K)^{-\epsilon}$ is generated by the non-zero vector $\delta(P_K)$ so that

$$\dim_{\mathbb{F}_p}(\mathrm{Sel}_p(E/K)^{-\epsilon}) = 1.$$

The results obtained in steps 1 and 2 imply that $\mathrm{Sel}_p(E/K)$ is a one-dimensional $\mathbb{F}_p$-vector space generated by $\delta(P_K)$. This completes the proof of theorem 10.2. $\square$

References

Kolyvagin's theorem is proved in [**Kol88**] and [**Kol89**]. Our exposition of Kolyvagin's method was influenced by the presentation in in [**Gr89**] and by the point of view taken in [**DDT95**] and [**Da02**]. Other useful references to the general machinery of Euler systems and Kolyvagin's argument can be found in [**Kol90**] and [**Ru00**] for example.

Exercises

(1) Suppose that p is a descent prime satisfying the hypotheses of chapter 10. Show that if L is any solvable extension of K then $E_p(L)$ is trivial.

(2) Suppose that p is a descent prime satisfying the hypotheses of chapter 10. Show that if $L = K(E_p)$, the restriction map

$$H^1(K, E_p) \longrightarrow H^1(L, E_p) = \hom(G_L, E_p)$$

is injective.

(3) Let K_v denote the completion of a number field K at a place v and let K_n^{nr} denote its maximal unramified extension. Suppose that v is a good place for M.
   (a) Show that $H^2(K_v^{\mathrm{nr}}/K_v, (K_v^{\mathrm{nr}})^\times)$ is trivial. Conclude that if c_1 and c_2 belong to $H^1_f(K_v, M)$ and $H^1_f(K_v, M^*)$ respectively, then
   $$\langle c_1, c_2\rangle_v = 0.$$
   (b) Show that v is a good place for M^*.
   (c) Let $\sigma_v \in \mathrm{Gal}(K_v^{\mathrm{nr}}/K_v)$ be the Frobenius element at v, and let $m = \#M$. Explicitly produce (*without* using the Tate local pairing) canonical identifications
   $$H^1_f(K_v, M) = M/(\sigma_v - 1)M, \qquad H^1_s(K_v, M^*) = \hom(\mu_m, M)[\sigma_v - 1].$$
   Conclude that $H^1_f(K_v, M)$ and $H^1_s(K_v, M^*)$ are in natural duality, so that in particular they have the same cardinality.

(d) Assuming the non-degeneracy of the local Tate pairing between $H^1(K_v, M)$ and $H^1(K_v, M)$, complete the proof of theorem 10.9 when v is a good prime for M.

(4) Let E be an elliptic curve over $\mathbb{Q}$ and let K be a quadratic extension of $\mathbb{Q}$. Let p be a descent prime satisfying the condtions that were imposed in the introduction of chapter 10.

(a) Show that the exension $L = K(E_{p^2})$ is Galois over $\mathbb{Q}$ with Galois group

$$\mathrm{Gal}(L/\mathbb{Q}) = \mathrm{Gal}(K/\mathbb{Q}) \times \mathrm{Gal}(\mathbb{Q}(E_{p^2})/\mathbb{Q}) = \mathrm{Gal}(K/\mathbb{Q}) \times \mathrm{Aut}(E_{p^2}),$$

so that elements in $\mathrm{Gal}(L/\mathbb{Q})$ are identified with pairs (τ^j, T) with $j \in \{0, 1\}$ and $T \in \mathbf{GL}_2(\mathbb{Z}/p^2\mathbb{Z})$.

(b) Show that there are infinitely many primes ℓ satisfying conditions (1) to (3) in definition 10.17 of a Kolyvagin prime.

(c) Suppose that K is a real quadratic field, and let u_K be a fundamental unit for K. Show that the extension $M = L(u_k^{1/p^2})$ obtained by adjoining to M a p^2-th root of u_K is Galois over $\mathbb{Q}$. Show that its Galois group can be described as the semi-direct product

$$\mathrm{Gal}(M/\mathbb{Q}) = \mu_{p^2} \rtimes \mathrm{Gal}(L/\mathbb{Q}),$$

where the action of $\mathrm{Aut}(E_{p^2})$ on μ_p^2 is the natural one, and the generator τ of $\mathrm{Gal}(K/\mathbb{Q})$ is made to act on μ_{p^2} as -1. Thus elements of $\mathrm{Gal}(M/\mathbb{Q})$ can be indexed by triples (ζ, τ^j, T) with ζ a p^2-th root of unity, $j \in \{0, 1\}$, and $T \in \mathrm{Aut}(E_{p^2}) \simeq \mathbf{GL}_2(\mathbb{Z}/p^2\mathbb{Z})$.

(d) Let $T \in \mathbf{GL}_2(\mathbb{Z}/p^2\mathbb{Z})$ be a matrix with eigenvalues a and $-1/a$, where $a \equiv 1 \pmod p$ but not modulo p^2. Show that any rational prime ℓ which is unramified in $M/\mathbb{Q}$ and whose frobenius element $\mathrm{Frob}_\ell(M/\mathbb{Q})$ satisfies

$$\mathrm{Frob}_\ell(M/\mathbb{Q}) = (1, -1, T)$$

is a Kolyvagin prime. Conclude that there are infinitely many Kolyvagin primes.

(e) Let s be any non-zero element of $H^1(K, E_p)$. Assume that s belongs to a specific eigenspace for the action of $\tau \in \mathrm{Gal}(K/\mathbb{Q})$, so that

$$\tau s = \delta s, \text{ for some } \delta \in \{1, -1\}.$$

Let M_s be the extension of M cut out by the image $\bar{s}$ of s under restriction to $H^1(M, E_p) = \mathrm{Hom}(\mathrm{Gal}(\bar{M}/M), E_p)$. Show that the extension M_s is Galois over $\mathbb{Q}$, and that $\mathrm{Gal}(M_s/\mathbb{Q})$ is identified with the semi-direct product

$$\mathrm{Gal}(M_s/\mathbb{Q}) = E_p \rtimes \mathrm{Gal}(M/\mathbb{Q}),$$

where the quotient $\mathrm{Gal}(M/\mathbb{Q})$ acts on the abelian normal subgroup E_p of $\mathrm{Gal}(M_s/\mathbb{Q})$ by the rule

$$(\zeta, \tau^j, T)(v) = \delta^j \bar{T} v.$$

(Here $\bar{T}$ denotes the natural image of T in $\mathrm{Aut}(E_p)$.)

(f) Show that the group $\mathrm{Gal}(M_s/\mathbb{Q})$ contains an element of the form $(v, 1, \tau, T)$, where the automorphism T is as in (d), and the vector $v \in E_p$ is non-zero and belongs to the δ-eigenspace for $\bar{T}$.

(g) Let $\ell \nmid N$ be a rational prime which is unramified in $M_s/\mathbb{Q}$ and satisfies

$$\mathrm{Frob}_\ell(M_s/\mathbb{Q}) = (v, 1, \tau, T).$$

Show that ℓ is a Kolyvagin prime, and that, if λ is the (unique) prime of K above ℓ, we have $s_\lambda \neq 0$. Conclude that there exist infinitely many Kolyvagin primes ℓ such that $\partial_\lambda(s) = 0$ and $s_\lambda \neq 0$.

(h) Let H be any finite-dimensional subspace of $H^1(K, E_p)$. Using (g), show that there is a finite set S of Kolyvagin primes with the property that the natural map induced by restriction

$$H \longrightarrow \oplus_v H^1(K_v, E_p)$$

is injective. Conclude proposition 10.18.

(i) Prove proposition 10.23.

(5) Following the notations in the proof of proposition 10.21, show that:

$$\tau D_\ell = -D_\ell \tau \pmod{n_\ell}, \quad \tau D_{\ell_1} D_{\ell_2} = D_{\ell_1} D_{\ell_2} \tau \pmod{\gcd(n_{\ell_1}, n_{\ell_2})}.$$

(In particular, these identities hold modulo p^2.)

(6) Give the details of the construction of the cohomology classes $\kappa(\ell_1 \ell_2)$ depending on two Kolyvagin primes, and prove the properties asserted in proposition 10.22.

Bibliography

[AL70] A.O.L. Atkin and J. Lehner. *Hecke operators on $\Gamma_0(m)$.* Math. Ann. **185** (1970) 134–160.

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor. *On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises.* J. Amer. Math. Soc. **14** (2001), no. 4, 843–939.

[BC92] J-F. Boutot and H. Carayol. *Uniformisation p-adique des courbes de Shimura: les théorèmes de Cerednik et de Drinfeld.* Courbes modulaires et courbes de Shimura (Orsay, 1987/1988). Astérisque No. 196-197 (1991) 45–158.

[BD96] M. Bertolini and H. Darmon. *Heegner points on Mumford-Tate curves.* Invent. Math **126** (1996) 413–456.

[BD97] M. Bertolini and H. Darmon. *A rigid-analytic Gross-Zagier formula and arithmetic applications.* Annals of Math **146** (1997) 111-147.

[BD98] M. Bertolini and H. Darmon. *Heegner points, p-adic L-functions, and the Cerednik-Drinfeld uniformisation.* Invent. Math. **131** (1998) 453–491.

[BD99] M. Bertolini and H. Darmon. *p-adic periods, p-adic L-functions and the p-adic uniformisation of Shimura curves.* Duke Math. J. **98** (1999), no. 2, 305–334.

[BD01] M. Bertolini and H. Darmon. *The p-adic L-functions of modular elliptic curves.* Mathematics unlimited—2001 and Beyond, 109–170, Springer-Verlag, Berlin, 2001.

[BD03] M. Bertolini and H. Darmon. *Iwasawa's main conjecture for elliptic curves in the anticyclotomic setting.* Annals of Mathematics, to appear.

[BDIS02] M. Bertolini, H. Darmon, A. Iovita, and M. Spiess. *Teitelbaum's conjecture in the anticyclotomic setting.* American Journal of Mathematics **124** (2002) 411-449.

[BDG03] M. Bertolini, H. Darmon, and P. Green. *Periods and points attached to quadratic algebras.* To appear in the Proceedings of an MSRI workshop on special values of Rankin L-series. H. Darmon and S. Zhang, eds.

[BFH90] D. Bump, S. Friedberg and J. Hoffstein. *Eisenstein series on the metaplectic group and nonvanishing theorems for automorphic L-functions and their derivatives.* Ann. of Math. (2) **131** (1990), no. 1, 53–127.

[Br94] K.S. Brown. Cohomology of groups. Corrected reprint of the 1982 original. Graduate Texts in Mathematics **87** Springer-Verlag, New York, 1994.

[BSD63] B.J. Birch and H.P.F. Swinnerton-Dyer. Notes on elliptic curves. I. J. Reine Angew. Math. **212** (1963) 7–25.

[BSD65] B.J. Birch and H.P.F. Swinnerton-Dyer. Notes on elliptic curves. II. J. Reine Angew. Math. **218** (1965) 79–108.

[Bu97] D. Bump. Automorphic forms and representations. Cambridge Studies in Advanced Mathematics, **55**. Cambridge University Press, Cambridge, 1997.

[Car86] H. Carayol. *Sur la mauvaise réduction des courbes de Shimura.* Compositio Math. **59** (1986), no. 2, 151–230.

[Car91] H. Carayol. *Formes modulaires et représentations galoisiennes à valeurs dans un anneau local complet.* In p-adic monodromy and the Birch and Swinnerton-Dyer conjecture (Boston, MA, 1991), 213–237, Contemp. Math., 165, Amer. Math. Soc., Providence, RI, 1994.

[Cas91] J.W.S. Cassels. Lectures on elliptic curves. London Mathematical Society Student Texts, **24**. Cambridge University Press, Cambridge, 1991.

[CDT99] B. Conrad, F. Diamond, and R. Taylor. *Modularity of certain potentially Barsotti-Tate Galois representations.* J. Amer. Math. Soc. **12** (1999), no. 2, 521–567.

[Ce76] I.V. Čerednik. *Uniformization of algebraic curves by discrete arithmetic subgroups of $\mathrm{PGL}_2(k_w)$ with compact quotient spaces.* Mat. Sb. (N.S.) 100(142) (1976), no. 1, 59–88.

[CF67] Algebraic number theory. Proceedings of an instructional conference organized by the London Mathematical Society. Edited by J. W. S. Cassels and A. Frhlich. Academic Press, London; Thompson Book Co., Inc., Washington, D.C. 1967.

[Con] K. Conrad. *On a conjecture of Birch and Swinnerton-Dyer: Partial Euler products on the critical line.* Canadian Journal of Mathematics, to appear.

[Cor02] C. Cornut. *Mazur's conjecture on higher Heegner points.* Invent. Math. **148** (2002), no. 3, 495–523.

[Cox89] D.A. Cox. Primes of the form $x^2 + ny^2$. Fermat, class field theory and complex multiplication. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1989.

[Cr97] J.E. Cremona. Algorithms for modular elliptic curves. Second edition. Cambridge University Press, Cambridge, 1997.

[Da92] H. Darmon. *Heegner points, Heegner cycles, and congruences.* "Elliptic curves and related topics", CRM proceedings and lecture notes vol. **4**, H. Kisilevsky and M. Ram Murty eds. (1992) 45–60.

[Da96] H. Darmon. *Stark-Heegner points over real quadratic fields.* Number theory (Tiruchirapalli, 1996), 41–69, Contemp. Math., **210**, Amer. Math. Soc., Providence, RI, 1998.

[Da01] H. Darmon. *Integration on $\mathcal{H}_p \times \mathcal{H}$ and arithmetic applications.* Annals of Mathematics **154** (2001) 589–639.

[Da02] H. Darmon. *Review of the book "Euler Systems" by Karl Rubin.* Bull. Amer. Math. Soc. **39** (2002) 407–414.

[Da03] H. Darmon. *Heegner points and elliptic curves of large rank over function fields.* To appear in the proceedings of an MSRI Workshop on Special Values of Rankin L-series. H. Darmon and S. Zhang, eds.

[DDT95] H. Darmon, F. Diamond, and R. Taylor. *Fermat's last theorem.* Current developments in mathematics, 1995 (Cambridge, MA), 1–154, Internat. Press, Cambridge, MA, 1994. Reprinted in: Elliptic curves, modular forms & Fermat's last theorem (Hong Kong, 1993), 2–140, Internat. Press, Cambridge, MA, 1997.

[DG01] H. Darmon and P. Green. *Elliptic curves and class fields of real quadratic fields: algorithms and evidence.* Journal of Experimental Mathematics **11:1** (2002) 37-55.

[DL03] H. Darmon and A. Logan. *Periods of Hilbert modular forms and rational points on elliptic curves.* International Mathematics Research Notices, submitted.

[Di96] F. Diamond. *On deformation rings and Hecke rings.* Ann. of Math. (2) **144** (1996) 137–166.

[DI95] F. Diamond and J. Im. *Modular forms and modular curves.* Seminar on Fermat's Last Theorem (Toronto, ON, 1993–1994), 39–133, CMS Conf. Proc., 17, Amer. Math. Soc., Providence, RI, 1995.

[Dr76] Drinfeld, V. G. *Coverings of p-adic symmetric domains.* Funkcional. Anal. i Prilozen. **10** (1976), no. 2, 29–40.

[DS95] E. de Shalit, *p-adic periods and modular symbols of elliptic curves of prime conductor.* Invent. Math. **121** (1995), no. 2, 225–255.

[Ed89] B. Edixhoven, *On the Manin constants of modular elliptic curves.* Arithmetic algebraic geometry (Texel, 1989), 25–39, Progr. Math., **89**, Birkhäuser Boston, Boston, MA, 1991.

[EGM98] J. Elstrodt, F. Grunewald, and J. Mennicke. Groups acting on hyperbolic space. Harmonic analysis and number theory. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998.

[Frei90] E. Freitag. Hilbert modular forms. Springer-Verlag, Berlin, 1990.

[Gar90] P.B. Garrett. Holomorphic Hilbert modular forms. The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990.

[Gel75] S.S. Gelbart. Automorphic forms on adle groups. Annals of Mathematics Studies, No. 83. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1975.

[GKZ87] B. Gross, W. Kohnen,and D. Zagier. *Heegner points and derivatives of L-series. II.* Math. Ann. **278** (1987), no. 1-4, 497–562.

[Go] D. Goldfeld. *Sur les produits partiels eulériens attachés aux courbes elliptiques.* C. R. Acad. Sci. Paris Sr. I Math. **294** (1982), no. 14, 471–474.

[Gr84] B.H. Gross. *Heegner points on* $X_0(N)$. Modular forms (Durham, 1983), 87–105, Ellis Horwood Ser. Math. Appl.: Statist. Oper. Res., Horwood, Chichester, 1984.

[Gr87] B.H. Gross, *Heights and the special values of L-series.* Number theory (Montreal, Que., 1985), 115–187, CMS Conf. Proc., **7**, Amer. Math. Soc., Providence, RI, 1987.

[Gr89] B.H. Gross. *Kolyvagin's work on modular elliptic curves.* L-functions and arithmetic (Durham, 1989), 235–256, London Math. Soc. Lecture Note Ser., 153, Cambridge Univ. Press, Cambridge, 1991.

[GvdP80] L. Gerritzen and M. van der Put. Schottky groups and Mumford curves. Lecture Notes in Mathematics, **817**. Springer, Berlin, 1980.

[GS93] R. Greenberg and G. Stevens. *p-adic L-functions and p-adic periods of modular forms.* Invent. Math. **111** (1993), no. 2, 407–447.

[GZ84] B.H. Gross and D.B. Zagier. *Heegner points and derivatives of L-series.* Invent. Math. **84** (1986), no. 2, 225–320.

[Ha75] G. Harder. *On the cohomology of discrete arithmetically defined groups.* Discrete subgroups of Lie groups and applications to moduli (Internat. Colloq., Bombay, 1973), pp. 129–160. Oxford Univ. Press, Bombay, 1975.

[HST93] M. Harris, D. Soudry, and R. Taylor. *l-adic representations associated to modular forms over imaginary quadratic fields. I. Lifting to* $\mathrm{GSp}_4(Q)$. Invent. Math. **112** (1993), no. 2, 377–411.

[Hu87] D. Husemoller. Elliptic curves. With an appendix by Ruth Lawrence. Graduate Texts in Mathematics, **111**. Springer-Verlag, New York, 1987.

[Ih68] Y. Ihara. *On congruence monodromy problems.* Vol. 1. Lecture Notes, No. 1 Department of Mathematics, University of Tokyo, Tokyo 1968.

[Ih79] Y. Ihara. *Congruence relations and Shimura curves.* Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2, pp. 291–311, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979.

[Ja72] H. Jacquet. *Automorphic forms on* GL(2). *Part II.* Lecture Notes in Mathematics **278**. Springer-Verlag, Berlin-New York, 1972.

[JL70] H. Jacquet and R.P. Langlands. Automorphic forms on GL(2). Lecture Notes in Mathematics **114**. Springer-Verlag, Berlin-New York, 1970.

[Ka92] S. Katok. Fuchsian groups. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1992.

[Kl91] C. Klingenberg, *On p-adic L-functions of Mumford curves.* p-adic monodromy and the Birch and Swinnerton-Dyer conjecture (Boston, MA, 1991), 277–315, Contemp. Math., **165**, Amer. Math. Soc., Providence, RI, 1994.

[Kn92] A.W. Knapp. Elliptic curves. Mathematical Notes, 40. Princeton University Press, Princeton, NJ, 1992.

[Kob93] N. Koblitz. Introduction to elliptic curves and modular forms. Second edition. Graduate Texts in Mathematics **97**. Springer-Verlag, New York, 1993.

[Kol88] V.A. Kolyvagin. *Finiteness of* $E(Q)$ *and* $\underline{III}(E,\mathbb{Q})$ *for a subclass of Weil curves.* (Russian) Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 3, 522–540, 670–671; translation in Math. USSR-Izv. **32** (1989), no. 3, 523–541.

[Kol89] V.A. Kolyvagin. *The Mordell-Weil and Shafarevich-Tate groups for Weil elliptic curves.* (Russian) Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 6, 1154–1180, 1327; translation in Math. USSR-Izv. **33** (1989), no. 3, 473–499.

[Kol90] V.A. Kolyvagin. Euler systems. The Grothendieck Festschrift, Vol. II, 435–483, Progr. Math. **87**, Birkhäuser Boston, Boston, MA, 1990.

[Men67] J. Mennicke, *On Ihara's modular group.* Invent. Math. **4** (1967) 202–228.

[Mes91] J.-F. Mestre. *Courbes elliptiques de rang* ≥ 12 *sur* $Q(t)$. C. R. Acad. Sci. Paris Sr. I Math **313** (1991), no. 4, 171–174.

[Mi86] J.S. Milne. Arithmetic duality theorems. Perspectives in Mathematics, 1. Academic Press, Inc., Boston, MA, 1986.

[MM91] M.R. Murty and V.K. Murty. *Mean values of derivatives of modular L-series.* Ann. of Math. (2) **133** (1991), no. 3, 447–475.

[MM97] M.R. Murty and V.K. Murty. *Non-vanishing of L-functions and applications.* Progress in Mathematics **157**. Birkhuser Verlag, Basel, 1997.

[MTT] B. Mazur, J. Tate, and J. Teitelbaum. *On p-adic analogues of the conjectures of Birch and Swinnerton-Dyer.* Invent. Math. **84** (1986), no. 1, 1–48.

[MSw-D74] B. Mazur and P. Swinnerton-Dyer. *Arithmetic of Weil curves.* Invent. Math. **25** (1974) 1–61.

[MS78] Y. Matsushima and G. Shimura. *On the cohomology groups attached to certain vector-valued differential forms on the product of upper half planes.* Ann. of Math. (2) **78** (1963) 417–449.

[Mu] V.K. Murty. Introduction to abelian varieties. CRM Monograph Series, 3. American Mathematical Society, Providence, RI, 1993.

[Od82] T. Oda. Periods of Hilbert modular surfaces. Progress in Mathematics **19**. Birkhäuser, Boston, Mass., 1982.

[Og69] A. Ogg. Modular forms and Dirichlet series. W. A. Benjamin, Inc., New York-Amsterdam 1969.

[Ri94] K.A. Ribet. *Fields of definition of abelian varieties with real multiplication.* Arithmetic geometry (Tempe, AZ, 1993), 107–118, Contemp. Math., 174, Amer. Math. Soc., Providence, RI, 1994.

[Ro96] D.E. Rohrlich. *Galois theory, elliptic curves, and root numbers.* Compositio Math. **100** (1996), no. 3, 311–349.

[Ru00] K. Rubin. Euler systems. Annals of Mathematics Studies **147**. Hermann Weyl Lectures. The Institute for Advanced Study. Princeton University Press, Princeton, NJ, 2000.

[Sch84] P. Schneider. *Rigid-analytic L-transforms.* Number theory, Noordwijkerhout 1983 (Noordwijkerhout, 1983), 216–230, Lecture Notes in Math **1068**, Springer, Berlin, 1984.

[Se67] J.-P. Serre. *Complex multiplication.* 1967 Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965) pp. 292–296 Thompson, Washington, D.C.

[Se70] J.-P. Serre. *Le problème des groupes de congruence pour* $\mathbf{SL}_2$. Ann. of Math. (2) **92** (1970) 489–527.

[Se71] J.-P. Serre. *Cohomologie des groupes discrets.* Prospects in mathematics (Proc. Sympos., Princeton Univ., Princeton, N.J., 1970), pp. 77–169. Ann. of Math. Studies, No. 70, Princeton Univ. Press, Princeton, N.J., 1971.

[Se72] J.-P. Serre. *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques.* Invent. Math. **15** (1972), no. 4, 259–331.

[Se80] J.-P. Serre. Trees. Translated from the French by John Stillwell. Springer-Verlag, Berlin-New York, 1980.

[Sh64] G. Shimura, *Class-fields and automorphic functions.* Ann. of Math. (2) **80** (1964) 444–463.

[Sh67] G. Shimura, *Construction of class fields and zeta functions of algebraic curves.* Ann. of Math. (2) **85** (1967) 58–159.

[Sh71] G. Shimura. Introduction to the arithmetic theory of automorphic functions. Reprint of the 1971 original. Publications of the Mathematical Society of Japan, 11. Kanô Memorial Lectures, 1. Princeton University Press, Princeton, NJ, 1994.

[Sh86] G. Shimura, *Algebraic number fields and symplectic discontinuous groups.* Ann. of Math. (2) **86** (1967) 503–592.

[Si86] J.H. Silverman. The arithmetic of elliptic curves. Corrected reprint of the 1986 original. Graduate Texts in Mathematics **106**. Springer-Verlag, New York, 1986.

[Si94] J.H. Silverman. Advanced topics in the arithmetic of elliptic curves. Graduate Texts in Mathematics **151**. Springer-Verlag, New York, 1994.

[Sta87] H.M. Stark. *Modular forms and related objects.* Number theory (Montreal, Que., 1985), 421–455, CMS Conf. Proc., 7, Amer. Math. Soc., Providence, RI, 1987.

[ST92] J.H. Silverman and J. Tate. *Rational points on elliptic curves.* Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1992.

[Ta62] J. Tate. *Duality theorems in Galois cohomology over number fields.* 1963 Proc. Internat. Congr. Mathematicians (Stockholm, 1962) pp. 288–295, Inst. Mittag-Leffler, Djursholm.

[Ta72] J. Tate. *Algorithm for determining the type of a singular fiber in an elliptic pencil.* Modular functions of one variable, IV (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), pp. 33–52. Lecture Notes in Math. **476**, Springer, Berlin, 1975.

[Ta74] J. Tate. *The arithmetic of elliptic curves.* Invent. Math. **23** (1974) 179–206.

[Te90] J.T. Teitelbaum. *Values of p-adic L-functions and a p-adic Poisson kernel.* Invent. Math. **101** (1990), no. 2, 395–410.

[TS67] J.T. Tate and I.R. Šafarevič. *The rank of elliptic curves.* (Russian) Dokl. Akad. Nauk SSSR **175** (1967) 770–773.

[T94] R. Taylor. *l-adic representations associated to modular forms over imaginary quadratic fields. II.* Invent. Math. **116** (1994), no. 1-3, 619–643.

[TW95] R. Taylor and A. Wiles. *Ring-theoretic properties of certain Hecke algebras.* Ann. of Math. (2) **141** (1995), no. 3, 553–572.

[Ul] D. Ulmer. *Elliptic curves with large rank over function fields.* Ann. of Math. (2) **155** (2002), no. 1, 295–315.

[Vi80] M.-F. Vignéras. Arithmétique des algèbres de quaternions. Lecture Notes in Mathematics **800**. Springer, Berlin, 1980.

[Va02] V. Vatsal. *Uniform distribution of Heegner points.* Invent. Math. **148** (2002), no. 1, 1–46.

[Wa85] J.-L. Waldspurger. *Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie.* Compositio Math. **54** (1985), no. 2, 173–242.

[Wi95] A. Wiles. *Modular elliptic curves and Fermat's last theorem.* Ann. of Math. (2) **141** (1995), no. 3, 443–551.

[Wi00] A. Wiles. *The Birch and Swinnerton-Dyer Conjecture*, Clay Mathematics Institute Web Site, http://www.claymath.org/prizeproblems/birchsd.pdf.

[Za85] D. Zagier, D. *Modular points, modular curves, modular surfaces and modular forms.* Workshop Bonn 1984 (Bonn, 1984), 225–248, Lecture Notes in Math. **1111**, Springer, Berlin, 1985.

[Zh01a] S. Zhang. *Heights of Heegner points on Shimura curves.* Ann. of Math. (2) **153** (2001), no. 1, 27–147.

[Zh01b] S. Zhang. *Gross-Zagier formula for* GL_2. Asian J. Math. **5** (2001), no. 2, 183–290.