MATH 596: Topics in Algebra and Number Theory

Jan Vonk
McGill University, Winter 2017

Contents

Introduction

These are the notes for a topics course in algebra and number theory, taught at McGill University from January – April 2017. The narrative is loosely centered around the topic of complex multiplication for elliptic curves, and is divided in three chapters. The intended audience is a group of advanced undergraduate and beginning graduate students, most of whom attended a course on modular forms and modular curves taught by Prof. Darmon the previous term.

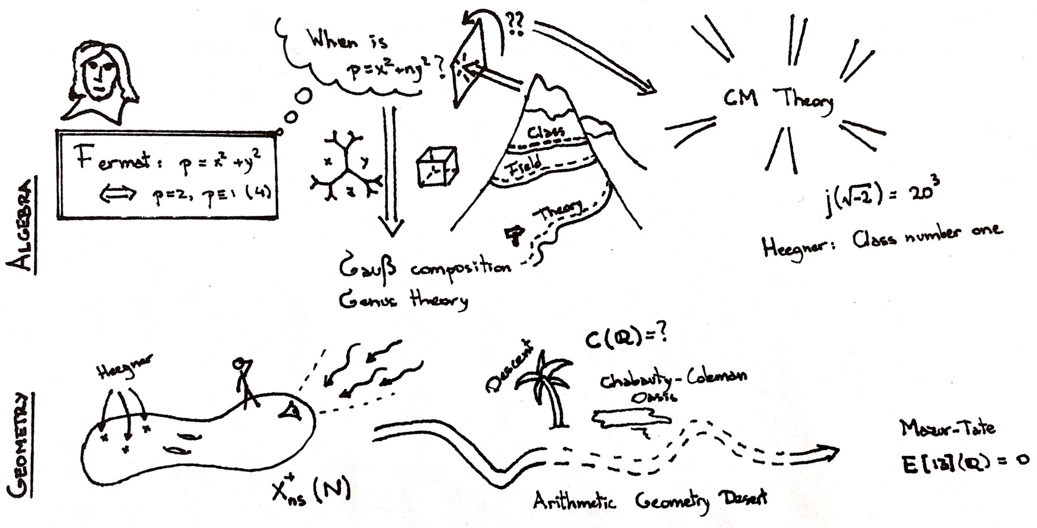


Figure 1: Leitfaden

We start with Fermat's result on primes which are the sum of two squares, for which we present three proofs due to Euler, Dedekind, and Heath–Brown/Zagier. We then consider the following generalisation

Question 1: Given $n \in \mathbf{Z}$, can we characterise primes of the form $x^2 + ny^2$?

The first chapter is concerned with answering this question, and we largely follow the excellent exposition of Cox [Cox89], the only notable difference being that we do not restrict to positive definite forms, and systematically attempt to include indefinite forms into the discussion. We discuss quadratic forms and Gauß composition from a modern perspective using Conway's topograph and Bhargava cubes, and finally come to a theoretical resolution of the above question at the end of the first part, using class field theory. Our treatment of global class field theory is entirely utilitarian, and does not provide proofs of the main theorems.

We then ask ourselves how much more explicit we can make the criteria for a prime to be represented by $x^2 + ny^2$. This is equivalent to the question of how to find generators for ring class fields explicitly. When n is positive, the theory of complex multiplication for elliptic curves provides us with an algorithmic way to answer this question. We follow the expositions by Silverman [Sil09] and Shimura [Shi70], give three proofs of the integrality of the j-invariant of a CM elliptic curve, and present the main theorem of complex multiplication. These theoretical tools are then put into practice by calculating the value of j on some examples using the Weber functions, and we then present Heegner's original answer to the following question.

Question 2: How many quadratic imaginary number fields have class number one?

As we will see, there are exactly 9 such fields. The theory of complex multiplication gives a beautiful way to describe abelian extensions of quadratic imaginary number fields, and the natural question arises whether this may be generalised to other number fields. This question was first raised by Leopold Kronecker (7 December 1823 – 29 December 1891) and is more commonly known as *Kronecker's Jugendtraum*, or nowadays, as Hilbert's 12th problem. We discuss some recent work of Duke–Imamoglu–Tóth on the case of real quadratic fields.

In the course of Heegner's proof of the class number one problem, a number of auxiliary Diophantine equations needed to be solved. This makes us wonder whether we can give a clean geometric interpretation, and we will discuss the reformulation due to Serre in terms of finding rational points on *non-split Cartan* modular curves. We then find ourselves confronted with the question of finding explicitly all rational points on curves defined over $\mathbf{Q}$, and discuss descent via isogeny on abelian varieties, as well as the method of Chabauty–Coleman for carrying out this process on concrete examples. We discuss various ways of investigating rational points on modular curves, and naturally obtain a strategy to address the following question.

Question 3: What possible orders can a rational torsion point on an elliptic curve $E_{\mathbf{Q}}$ have?

We finish with a result due to Mazur–Tate that assures us that no rational 13-torsion point can exist. This of course is a toy case of Mazur's torsion theorem, which we will keep for next time.

Acknowledgements. I would like to thank Joachim De Ronde, Michele Fornea, David Ter Borch Gram Lillienfeldt, Forte Shinko, Nicolas Simard for pointing out numerous errors in the text, and making useful suggestions.

Chapter I

Binary Quadratic Forms

1 Musings of Fermat and Euler

We begin by recalling a classical result due to Fermat, which states that a prime with remainder one upon division by four is always the sum of two squares. We give three proofs due to Euler, Dedekind, and Heath–Brown, and present some more general conjectures due to Euler. These conjectures will be our motivation for developing the theory of quadratic forms in this chapter.

1.1 Fermat's conjectures.

Fermat conjectured that an odd prime p is of the form $x^2 + y^2$ for integers x, y if and only if $4 \mid p - 1$. As he has been known to do, he claimed to have a proof but did not provide one. Likewise, he conjectured that

$$\begin{cases} p = x^2 + 2y^2 \quad \iff \quad p \equiv 1, 3 \pmod 8 \\ p = x^2 + 3y^2 \quad \iff \quad p \equiv 1 \pmod 3 \end{cases}$$

Once more, Fermat claimed to have *firmissimis demonstratibus* but did not give any details about them, other than that they were related to his famous technique of descent. It was not until about a century later that Euler became interested in these questions, and provided full proofs for these claims. In the terminology of Cox [Cox89], the main steps in Euler's argument may be labelled as follows:

1. Reciprocity step: $p \equiv 1 \pmod 4 \Rightarrow p \mid a^2 + b^2$, some $a, b \in \mathbf{Z}$
2. Descent step: $p \mid a^2 + b^2 \Rightarrow p = c^2 + d^2$, some $c, d \in \mathbf{Z}$

1.2 Euler's proof.

As discussed above, there are two main parts to Euler's proof. The *reciprocity step* is the easiest. It is a well-known fact that -1 is a square modulo an odd prime p if and only if $p \equiv 1 \pmod 4$. In other words,

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} \tag{I.1}$$

Indeed, if $\alpha^2 \equiv -1 \pmod p$ then $(-1)^{(p-1)/2} \equiv \alpha^{p-1} \equiv 1 \pmod p$. As p is odd, this means that $(-1)^{(p-1)/2} = 1$ and hence $(p-1)/2 = 2k$ even, giving us $p = 4k+1$. Conversely, if $p = 4k+1$ then the polynomial

$$x^{4k} - 1 = (x^{2k} - 1)(x^{2k} + 1)$$

has $4k$ roots in $\mathbf{F}_p^\times$ by Fermat's little theorem, and as $\mathbf{F}_p$ is a field this means $(x^{2k}+1)$ must have exactly $2k$ roots. Any root α then satisfies $(\alpha^k)^2 \equiv -1 \pmod p$ so -1 is a square. In conclusion, we can always find an $\alpha \in \mathbf{Z}$ such that $p \mid \alpha^2 + 1$.

The argument of Euler for the *descent step* is as follows. It starts with the identity

$$(x^2+y^2)(z^2+w^2) = (xz+yw)^2 + (xw-yz)^2. \tag{I.2}$$

Now suppose that $N = a^2+b^2$ with $\gcd(a,b) = 1$, and $q = x^2+y^2$ a prime divisor of N. Then $q \mid x^2N - a^2q = (xb-ay)(xb+ay)$ and hence without loss of generality we deduce $qd = xb - ay$ for some $d \in \mathbf{Z}$. Now, we have $x \mid xb - dx^2 = (a+dy)y$ and as x, y are coprime we also have $a + dy = cx$ for some $c \in \mathbf{Z}$. From this, we quickly deduce that

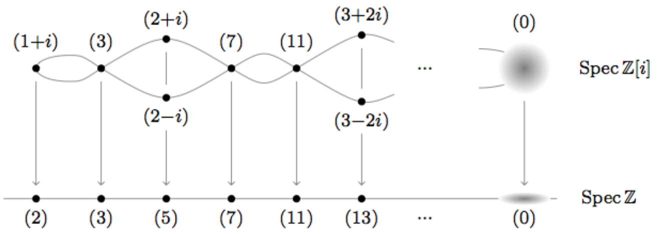
$$\begin{cases} a &= cx - dy \\ b &= dx + cy \end{cases}$$

The identity (I.2) now implies that $N = q(c^2+d^2)$ and hence that $\frac{N}{q} = c^2 + d^2$ is also the sum of two squares.

To complete the descent step, just note that if $p \mid a^2 + b^2$ then we may assume that $2|a|, 2|b| < p$ and a, b coprime. It follows that $N = a^2 + b^2 < p^2/2$, and hence any other prime factor q of N must be less than p. As we have just seen, if q is the sum of two squares, then also N/q is the sum of two squares. We now argue by induction on the size of p that p must also be the sum of two squares.

1.3 Dedekind's proof.

Dedekind had studied the splitting behaviour of ideals in extensions of number fields. His proof is clean and short, though it uses the correspondence between splitting of ideals and splittings of polynomials. Note that it also makes crucial use of the fact that -1 is a square whenever $p \equiv 1 \pmod 4$.



The proof takes place in the ring of integers of $\mathbf{Q}(i)$, which is

$$\mathbf{Z}[i] \simeq \mathbf{Z}[x]/(x^2+1).$$

By Kummer–Dedekind, which is recalled in Appendix C, the factorisation of the ideal (p) in $\mathbf{Z}[i]$ for a prime $p \neq 2$ can be computed via the factorisation of the polynomial

$$x^2+1 \pmod{p}.$$

We know that -1 is a square modulo p if and only if $p \equiv 1 \pmod 4$, and so (p) remains prime in $\mathbf{Z}[i]$ if and only if $p \equiv 3 \pmod 4$ and splits completely otherwise. As $\mathbf{Z}[i]$ is a principal ideal domain, we obtain that

$$(p) = (a+bi)(a-bi), \quad a, b \in \mathbf{Z}, \text{ whenever } p \equiv 1 \pmod 4.$$

By multiplying out the generators of these ideal, we get that $i^n p = a^2+b^2$ for some n, but as the right hand side is a positive integer, we must necessarily have that $p = a^2+b^2$.

1.4 Heath-Brown's proof.

In 1971, Heath–Brown discovered a very short proof of the fact that any prime number $p = 4m+1$ is the sum of two squares. Zagier's simplified version of this proof starts by considering the set

$$S^+ = \{(x,y,z) \in \mathbf{N}^3 \ : \ 4yz + x^2 = p\}.$$

This is a finite set. Consider the involution on the set S^+ given by

$$(x,y,z) \mapsto \begin{cases} (x+2z, z, y-x-z) & \text{if} \quad x < y-z \\ (2y-x, y, x-y+z) & \text{if} \quad y-z < x < 2y \\ (x-2y, x-y+z, y) & \text{if} \quad 2y < x \end{cases}$$

It has exactly one fixed point $(1,1,m)$. Therefore, S^+ has an odd number of elements. This implies that the involution $(x,y,z) \mapsto (x,z,y)$ must also have a fixed point, and hence p is the sum of two squares. □

Whereas the reader can easily verify that the above assertions are all true, it is very hard to get a conceptual understanding of whence this proof came. Zagier's formulation above is extremely short, but Heath-Brown's original argument offers slightly more insight. It considered instead a slightly different set

$$S = \{(x,y,z) \in \mathbf{Z}^3 \ : \ 4xy + z^2 = p, \ \ x,y > 0\},$$

as well as the following three involutions on S:

$$\begin{array}{llll} \psi_1 : & S \to S, & (x,y,z) \mapsto (y,x,-z) \\ \psi_2 : & S \to S, & (x,y,z) \mapsto (x-y+z, y, 2y-z) \\ \psi_3 : & S \to S, & (x,y,z) \mapsto (y,x,z) \end{array}$$

Clearly, the first and last involution are rather trivial, and it is the second one that will capture some essential properties of the set S. Indeed, checking that it defines an involution on S uses the precise form of the definition of S, and it is clear that whatever the argument is, the involution ψ_2 will play the crucial role.

Here is how Heath-Brown puts these three involutions into action: Firstly, note $\psi_1(S^+) = S^-$, where S^+ and S^- are the sets of elements in S whose last coordinate is positive and negative, respectively. Note that the last coordinate is never zero, so $S^+ \cup S^- = S$. Likewise, if we define

$$T^+ = \{(x,y,z) \in S \ : \ x-y+z > 0\}, \ \ T^- = \{(x,y,z) \in S \ : \ x-y+z < 0\}$$

then it is easy to check that $\psi_1(T^+) = T^-$ and $S = T^+ \cup T^-$. It follows that ψ_1 interchanges the elements of $S^+ \backslash T^+$ and $T^+ \backslash S^+$, and so $|S^+| = |T^+|$. Secondly, it is easy to check that ψ_2 preserves T^+. In T^+, it has exactly one fixed point $(1,1,m)$. This means that T^+, and hence also S^+, must have an odd number of elements. Finally, as ψ_3 clearly preserves the set S^+, this implies that it must have a fixed point there, and hence p is the sum of two squares.

1.5 Particularities and generalities.

As well as proving Fermat's conjectures, Euler raised the more general question of when a prime p is of the form $x^2 + ny^2$ and treated some special cases. For instance, he conjectured

$$p = x^2 + 5y^2 \iff p \equiv 1, 9 \pmod{20},$$

but was unable to prove it using the same techniques as above. Furthermore, for larger values of n the criterion seems to get less precise, and not entirely determined by congruence conditions on the prime. An interesting example is given by another of Euler's conjectures:

$$\left\{ \begin{array}{lcl} p & = & x^2 + 14y^2, \text{ or} \\ p & = & 2x^2 + 7y^2 \end{array} \right. \iff p \equiv 1, 9, 15, 23, 25, 39 \pmod{56}.$$

Clearly, simple congruences do not seem to capture the full depth of when a prime is represented by a quadratic form. In some cases, Euler was able to formulate a conjecture which gives us a criterion that is significantly more sophisticated than a simple congruence. For example, he conjectures

$$p = x^2 + 27y^2 \iff \left\{ \begin{array}{l} p \equiv 1 \pmod{3} \text{ and} \\ t^3 - 2 \in \mathbf{F}_p[t] \text{ has a root.} \end{array} \right.$$

We can see something deep and interesting is going on, and will set out to determine which piece of mathematical theory explains and proved all of the above conjectures. Having the above three proofs of Fermat's theorem at our disposal, we wonder which one would be easiest to generalise to prove Euler's conjecture, or even address the general question of when a prime p is of the form $x^2 + ny^2$ for some given n.

Proof A. Euler's proof breaks up into two parts, and the *reciprocity step* is the generalisation of (I.1), and is clearly equivalent to determining whether $-n$ is a square modulo p. This question is one that we have studied many times before. It is rendered completely algorithmic by repeated applications of the following result.

Theorem 1.1 (Quadratic reciprocity). *If p, q are distinct odd primes, we have*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{(p-1)(q-1)}{4}},$$

and furthermore

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{(p-1)}{2}}, \qquad \left(\frac{2}{p}\right) = (-1)^{\frac{(p^2-1)}{8}}$$

Though something along those lines was known to other mathematicians, most credit for formulating, as well as finding the first four proofs for it, should go to Gauß, who discusses this *aureum theorema* in his *Disquisitiones Arithmeticae*. As it is assumed all of you have seen this theorem before, we will not prove it here, though various proofs are given 'en passant' later in these notes. The second *descent* step is very elementary and specific. Whereas the theory of Gauß compisition will put the Brahmagupta identity (I.2) in a general framework, the method of proof breaks down completely for general n. This is not surprising, as quadratic reciprocity is responsible for the *congruence* part of the above criteria, and congruences are not the end of the story as we saw in the more sophisticated criterion for $n = 27$.

Proof B. Dedekind's proof uses slightly more general theory about factorisation of prime ideals in extensions, but still relies heavily on the specifics of the case at hand. For instance, of crucial importance is the fact that $\mathbf{Z}[i]$ is a principal ideal domain, and if we try to mimic his argument in the more general rings $\mathbf{Z}[\sqrt{-n}]$ this is far too much to ask, these rings are in general not even Dedekind domains! However, it is perhaps the proof closest in spirit to what we will do when we finally solve the problem using ring class fields at the end of this chapter.

Proof C. Finally, I do not see how to generalise Heath–Brown's proof, but I encourage you to think about it and let me know what you find. Generalisations to other cases can be found in the literature, but I am unaware of a general framework that applies to a large class of cases.

2 Quadratic forms

In this section we formalise the notion of a quadratic form, and discuss various ways to study the natural action of $\mathrm{SL}_2(\mathbf{Z})$ on forms of a given discriminant. We discuss how to find distinguished elements in each orbit, the *"reduced forms"*, and investigate the stabilisers of forms, generated by *"automorphs"*. In subsequent sections, we will give a graphical interpretation of these concepts in terms of topographs. Excellent sources for the material in this section are Buell [Bue89] and Cox [Cox89].

2.1 Definitions.

A *binary quadratic form*, or simply quadratic form, is a homogenous polynomial of degree two in $\mathbf{Z}[x, y]$. We will usually write

$$F(x, y) = ax^2 + bxy + cy^2,$$

where $a, b, c \in \mathbf{Z}$ are the *coefficients* of F. The quadratic form F is said to be *primitive* when $\gcd(a, b, c) = 1$, and we will exclusively consider primitive forms in what follows. An important invariant of a quadratic form is its *discriminant* $\Delta = b^2 - 4ac$. Quadratic forms with $\Delta > 0$ are called *indefinite*, those with $\Delta < 0$ are called *definite*, and those with $\Delta = 0$ are called *parabolic*. Definite quadratic forms come in two flavours: Positive definite forms take only positive values, negative definite forms take only negative values. If there exist $x, y \in \mathbf{Z}$ such that $F(x, y) = n$, then we say that n is *represented* by F. If furthermore $x, y \in \mathbf{Z}$ may be chosen to be coprime, we say n is *properly represented*. Clearly, the conjectures of Fermat and Euler above are about the question of which prime numbers are represented by the quadratic forms

$$F_1(x, y) = x^2 + y^2, \qquad F_2(x, y) = x^2 + 5y^2, \qquad F_3(x, y) = x^2 + 14y^2$$

In this chapter, we will investigate representability of integers by quadratic forms. We will see that the above elementary arguments for F_1 are hardly sufficient for the general case, and we will see how the question takes us from the mathematics of Fermat, via the work of Gauß and others, to questions raised by Hilbert.

The central subject of study is the action of the group $\mathrm{SL}_2(\mathbf{Z})$, or occasionally the slightly larger group $\mathrm{GL}_2(\mathbf{Z})$, on the set of quadratic forms. This action is defined as follows. Let

$$\gamma = \begin{pmatrix} q & r \\ s & t \end{pmatrix} \in \mathrm{GL}_2(\mathbf{Z}), \qquad \text{where } \mathrm{GL}_2(\mathbf{Z}) = \left\{ \begin{pmatrix} q & r \\ s & t \end{pmatrix} \ : \ q, r, s, t \in \mathbf{Z},\ qt - rs = \pm 1 \right\}$$

then we define $F(x, y) \cdot \gamma = F(qx + ry, sx + ty)$. We can easily check that this defines a right action of $\mathrm{GL}_2(\mathbf{Z})$ on the set of quadratic forms which preserves primitive forms, and that the discriminant is invariant under this action. Commonly, we abbreviate $F(x, y) = ax^2 + bxy + cy^2$ as $\langle a, b, c\rangle$. In this notation, we can make the action of M on the coefficients explicit as follows:

$$\langle a, b, c\rangle \cdot \gamma = \langle A, B, C\rangle, \text{ where } \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} q^2 & qs & s^2 \\ 2qr & qt + rs & 2st \\ r^2 & rt & t^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Two quadratic forms in the same $\mathrm{SL}_2(\mathbf{Z})$-orbit are called *equivalent*, and two non-equivalent quadratic forms in the same $\mathrm{GL}_2(\mathbf{Z})$-orbit are called *improperly equivalent*. The group $\mathrm{SL}_2(\mathbf{Z})$ is generated by the elements

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

This gives rise to the equivalences $\langle a, b, c\rangle \sim \langle c, -b, a\rangle$ and $\langle a, b, c\rangle \sim \langle a, b + 2a, a + b + c\rangle$.

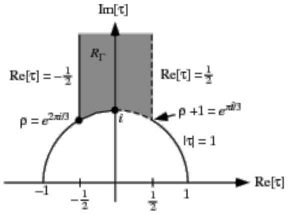
Forms equivalent under $\mathrm{GL}_2(\mathbf{Z})$ represent the same integers. Nonetheless, it is the slightly finer notion of $\mathrm{SL}_2(\mathbf{Z})$-equivalence that carries most interest, as was pointed out by Gauß. Given a discriminant Δ, can we describe the orbits of the $\mathrm{SL}_2(\mathbf{Z})$-action on the quadratic forms of that discriminant? Can we describe the stabiliser of a quadratic form? The answers for both questions are of a rather different nature for definite and indefinite forms, so we will treat those cases separately.

2.2 Definite quadratic forms.

For definite forms, every orbit consists of either positive or negative definite forms. The operation $\langle a, b, c\rangle \mapsto \langle -a, b, -c\rangle$ interchanges positive and negative definite forms, and there is no essential difference between the two theories. Let $F = \langle a, b, c\rangle$ be a primitive positive definite quadratic form. We say F is *reduced* if

$$|b| \leq a \leq c, \quad \text{and} \quad b \geq 0 \ \text{ if either } \ |b| = a \ \text{ or } \ a = c.$$

A negative definite form is reduced if its corresponding positive definite form is. Reduced forms will play the role of distinguished elements in an $\mathrm{SL}_2(\mathbf{Z})$-orbit. This definition has a more visual interpretation, as it is equivalent to saying that one of the roots of F lies in the standard fundamental domain for the action of $\mathrm{SL}_2(\mathbf{Z})$ via linear fractional transformation on the upper half plane $\mathfrak{H}$.



The usual proof that this is a fundamental domain, as discussed in Prof. Darmon's course or Serre [Ser70], shows that the following theorem is true:

Theorem 2.1. *Every definite form is equivalent to a unique reduced form.*

Proof. Let F be a definite form with root $\lambda \in \mathfrak{H}$, and $\gamma \in \mathrm{SL}_2(\mathbf{Z})$. Note that the root of $\gamma \cdot F$ in $\mathfrak{H}$ is $\gamma \cdot \lambda$, where the action is via linear fractional transformations. The proof that the above region is a fundamental domain therefore shows that any primitive $\mathrm{SL}_2(\mathbf{Z})$-orbit contains a unique reduced form. □

One immediate consequence of this theorem is that the number of equivalence classes of primitive forms for a given discriminant Δ is finite! Indeed, for a reduced form we have $b^2 \leq a^2$ and $a \leq c$, and hence $\Delta = b^2 - 4ac \leq a^2 - 4a^2 = -3a^2$. It follows that there are only finitely many possible values of a, and hence b, and hence also c. The number of primitive reduced forms of discriminant $\Delta < 0$ is called the *class number* h_Δ^+. This proof of finiteness also gives us a practical method to find all primitive reduced forms of a given discriminant $\Delta < 0$. Here is a table of some small values:

Δ	Reduced forms	Δ	Reduced forms
-3	$\pm\langle 1,1,1\rangle$	-19	$\pm\langle 1,1,5\rangle$
-4	$\pm\langle 1,0,1\rangle$	-20	$\pm\langle 1,0,5\rangle, \pm\langle 2,2,3\rangle$
-7	$\pm\langle 1,1,2\rangle$	-23	$\pm\langle 1,1,6\rangle, \pm\langle 2,\pm 1,3)$
-8	$\pm\langle 1,0,2\rangle$	-24	$\pm\langle 1,0,6\rangle, \pm\langle 2,0,3\rangle$
-11	$\pm\langle 1,1,3\rangle$	-27	$\pm\langle 1,1,7\rangle$
-12	$\pm\langle 1,0,3\rangle$	-28	$\pm\langle 1,0,7\rangle$
-15	$\pm\langle 1,1,4\rangle, \pm\langle 2,1,2\rangle$	-31	$\pm\langle 1,1,8\rangle, \pm\langle 2,\pm 1,4\rangle$
-16	$\pm\langle 1,0,4\rangle$	-32	$\pm\langle 1,0,8\rangle, \pm\langle 3,2,3\rangle$

If one continues this table further, one might conjecture that h_Δ^+ grows at the same rate as $\sqrt{-\Delta}$. Deuring, Heilbronn, and Siegel showed in the 1930's that for any positive constant $\epsilon > 0$, there exists some positive constant C_ϵ such that $h_\Delta^+ \geq C_\epsilon \cdot |\Delta|^{\frac{1}{2}-\epsilon}$, though this constant may not be computed explicitly. This implies that for any n there are at most finitely many discriminants $\Delta < 0$ such that $h_\Delta^+ = n$, though sadly it gives us no way find them. For $n = 1$, we will find all such discriminants in the next chapter. Note that if we wish to make this effective, the best we have is a result of Goldfeld from 1976, which shows that the existence of an elliptic curve $E_{\mathbf{Q}}$ whose L-series vanishes to order 3 at $s = 1$ implies that

$$h_\Delta^+ \geq C_E \cdot \log|\Delta|,$$

for all $\Delta < 0$. This is an amazing result, which has the advantage over Siegel's that C_E is effectively computable. All that remains is to find an elliptic curve whose L-series vanishes to order 3 at $s = 1$. We note that even the statement that the L-series of an elliptic curve has a continuation to $s = 1$ requires the modularity theorem. In 1983, Gross and Zagier found such an elliptic curve E, and Oesterlé determined we could take $C_E = 1/7000$. This gives us an effective way to find all discriminants with a given class number, though it is computationally extremely laborious and nowadays one might opt for different methods, or at least a better elliptic curve than that of Gross and Zagier. See Watkins [Wat03] for an extensive collection of tables of discriminants of class number up to 100, which were found using related methods.

Automorphisms. Suppose a matrix γ fixes a reduced quadratic form $F = \langle a,b,c\rangle$ of discriminant $\Delta < 0$, then it also fixes its root λ in the fundamental domain for the $\mathrm{SL}_2(\mathbf{Z})$-action on $\mathfrak{H}$, acting via linear fractional transformations. In previous courses on modular forms, you have called these *elliptic points*, and you know that there are precisely two such points, corresponding to the forms

$$\pm\langle 1,0,1\rangle, \qquad \text{and} \qquad \pm\langle 1,1,1\rangle.$$

As any definite quadratic form is equivalent to a unique reduced form, it follows that the stabiliser of any definite form is isomorphic to C_2, unless it is equivalent to one of the above two forms, in which case the stabiliser is isomorphic to C_4 and C_6 respectively.

2.3 Indefinite quadratic forms.

The theory for indefinite forms is significantly richer and more mysterious than its definite counterpart. A notion of reduced form that assures all the nice properties above is not known. Gauß introduced a notion of reduced form that still allows us to show that the number of orbits is finite. To simplify our presentation, we will always assume $\Delta > 0$ is not a square. The degenerate case where Δ is a square will be briefly discussed in the next section on Conway's topograph.

Following Gauß, we say that the indefinite form $F = \langle a, b, c\rangle$ of discriminant $\Delta > 0$ is *reduced* if

$$0 < \sqrt{\Delta} - b < 2|a| < \sqrt{\Delta} + b.$$

This condition is equivalent to the following condition on the roots $\lambda^- < \lambda^+$:

$$\begin{cases} \lambda^- < -1, \ \ \lambda^+ \in (0,1) & \text{if} \quad a \geq 0 \\ 1 < \lambda^+, \ \ \lambda^- \in (-1,0) & \text{if} \quad a < 0 \end{cases}$$

Again, reduced forms will play the role of distinguished elements in an $\mathrm{SL}_2(\mathbf{Z})$-orbit, with one very important difference: There can be many more than one reduced form per orbit! For instance, the two forms of discriminant $\Delta = 2021$ given by

$$\langle 5, 41, -17\rangle \qquad \text{and} \qquad \langle 19, 11, -25\rangle$$

are $\mathrm{SL}_2(\mathbf{Z})$-equivalent, even though they are both reduced.

Theorem 2.2. *Every indefinite form is equivalent to a reduced form.*

Proof. We say two forms $\langle a, b, c\rangle$ and $\langle c, d, e\rangle$ of discriminant $\Delta > 0$ are *neighbours* if $b + d \equiv 0 \pmod{2c}$. Then any two neighbours are equivalent by the transformation $ST^{(b+d)/2c} \in \mathrm{SL}_2(\mathbf{Z})$. Given a form $F = \langle a, b, c\rangle$, there is a canonical *right neighbour* $\rho(F) = \langle c, -b + 2sc, cs^2 - bs + a\rangle$ where

$$s = \begin{cases} \operatorname{sign}(c)\lfloor \frac{b}{2|c|} \rfloor & \text{if } |c| > \sqrt{\Delta} \\ \operatorname{sign}(c)\lfloor \frac{b+\sqrt{\Delta}}{2|c|} \rfloor & \text{if } |c| < \sqrt{\Delta} \end{cases}$$

Suppose that $|a| > \sqrt{\Delta}$, then $|c| < |a|/4$ and the size of the first coefficient of $\rho(F)$ is strictly less than that of F. By replacing F by $\rho^i(F)$ for some appropriate i, we may assume that $|a| < \sqrt{\Delta}$. An elementary check then shows that $\rho(F)$ is reduced. Therefore, by taking right neighbours repeatedly, we arrive at a reduced form in at most

$$\frac{1}{2}\log_2\left(\frac{|a|}{\sqrt{\Delta}}\right) + 2 \quad \text{steps.}$$

□

The above theorem implies that there are finitely many equivalence classes of forms of a given discriminant $\Delta > 0$, simply because there are only finitely many reduced forms of that discriminant. To see why, note that the inequalities imply $0 < b < \sqrt{\Delta}$ and therefore also $|a| < \sqrt{\Delta}$. For each of these possibilities, there are a finite number of possible values for c, and so at most finitely many reduced forms.

Cycles of reduced forms. The fact that we did not find a condition strong enough to cut out precisely one reduced form in every orbit may seem like an undesirable nuisance. However, a consideration of Conway's topograph below shows that this is in fact perfectly natural, and a reflection of what makes the theory of indefinite forms deeper and richer than the corresponding theory of definite forms.

In the above reduction algorithm, we continued to find right neighbours until we find a reduced form. Once this stage is reached, we could continue this process, and it is easy to see that every right neighbour will remain reduced. As the number of reduced forms is finite, this sequence will eventually become periodic. This periodic sequence is called a *cycle of reduced forms*. As the right neighbour has its first coefficient equal to the last coefficient of the previous form, we can denote a cycle consisting of the forms $\langle a_i, b_i, a_{i+1}\rangle$ by

$$a_0 \;{}^{b_0}\; a_1 \;{}^{b_1}\; a_2 \;{}^{b_2}\; a_3 \;\ldots$$

For example, if we take $F = x^2 - 67y^2$ which has $\Delta = 268$ then we obtain the cycle

$$+1\;{}^{0}\;-67\;{}^{0}\;+1\;{}^{16}\;-3\;{}^{14}\;+6\;{}^{10}\;-7\;{}^{4}\;+9\;{}^{14}\;-2\;{}^{14}\;+9\;{}^{4}\;-7\;{}^{10}\;+6\;{}^{14}\;-3\;{}^{16}\;+1\;{}^{16}\;-3\ldots$$

All forms from $\langle 1, 16, -3\rangle$ onwards are reduced, and the reduced cycle is periodic of length 10.

Theorem 2.3. *Two reduced indefinite quadratic forms are equivalent if and only if they are in the same cycle.*

Proof. See Buell [Bue89, Section 3.3]. □

Continued fractions. The ingredient that goes into the proof of the above theorem is the intimate connection between indefinite quadratic forms and the theory of continued fractions. We will have no immediate need for this in what follows, so we will just point out what can be read between the lines of the above reduction algorithm, and I will leave it to the reader to convince him/herself that really we are just doing continued fraction calculations. To get started, let us note that the continued fraction expansion of $\sqrt{67}$ is

$$\sqrt{67} = [8, \overline{5, 2, 1, 1, 7, 1, 1, 2, 5, 16}]$$

If we denote the convergents by p_n, q_n and we compute $c_n = p_n^2 - 67q_n^2$ and obtain the numbers

$$-3,\; 6,\; -7,\; 9,\; -2,\; 9,\; -7,\; 6,\; -3,\; 1,\; -3,\; 6,\; -7,\; 9,\; \ldots$$

Do you see the link with the reduced cycle? For a general indefinite quadratic form $F = \langle a, b, c\rangle$ we may compute the convergents p_n, q_n of the quadratic irrational $\frac{\sqrt{\Delta}-b}{2|a|}$ and set

$$c_n = a\, p_n^2 + \mathrm{sign}(a) b\, p_n q_n + c\, q_n^2.$$

These numbers form exactly the sequence of first coefficients of the reduced cycle attached to F.

Automorphisms. Suppose a matrix γ fixes a primitive quadratic form $F = \langle a, b, c\rangle$ of discriminant $\Delta < 0$, then it must also fix its roots. This gives rise to a quadratic polynomial with the same roots, which must therefore be proportional to the original quadratic form, yielding the equations

$$\begin{cases} ak &= t, \\ bk &= u-r, \\ ck &= -s. \end{cases} \qquad \text{where} \qquad \gamma = \begin{pmatrix} r & s \\ t & u \end{pmatrix},$$

for some $k \in \mathbf{Z}$. We deduce $(u+r)^2 - \Delta k^2 = 4$. Conversely, a solution to the *Pell equation* $T^2 - \Delta U^2 = 4$ gives us a matrix in the stabiliser of F, which therefore consists exactly of the matrices

$$\gamma = \pm \begin{pmatrix} \frac{T-bU}{2} & aU \\ -cU & \frac{T+bU}{2} \end{pmatrix}, \qquad \text{where} \qquad T^2 - \Delta U^2 = 4.$$

Let us now investigate whether we can always find solutions to the equation $T^2 - \Delta U^2 = 4$.

Theorem 2.4. *Let $\Delta > 0$ be a non-square discriminant. Then $T^2 - \Delta U^2 = 4$ always has a solution with $U \neq 0$. If we let (T_0, U_0) be the minimal solution with $T_0, U_0 > 0$ then every solution (X, Y) satisfies*

$$\frac{X + Y\sqrt{\Delta}}{2} = \pm \left(\frac{T_0 + U_0\sqrt{\Delta}}{2}\right)^n, \qquad \text{for some } n \in \mathbf{Z}.$$

Proof. Start with a quadratic form F of discriminant Δ, then the procedure of repeatedly taking right neighbours is eventually periodic, and hence we obtain non-trivial transformation matrices that fix any reduced form equivalent to F, giving us a solution to Pell's equation by the above discussion. To prove the second part, note that it is easy to check that $\frac{X+Y\sqrt{\Delta}}{2} = \pm\left(\frac{T_0+U_0\sqrt{\Delta}}{2}\right)^n$ always yields a solution, and given any positive solution (T, U) there must exist an $n \geq 1$ such that

$$\left(\frac{T_0 + U_0\sqrt{\Delta}}{2}\right)^n < \frac{T + U\sqrt{\Delta}}{2} \leq \left(\frac{T_0 + U_0\sqrt{\Delta}}{2}\right)^{n+1}.$$

Multiplying everything by $\left(\frac{T_0-U_0\sqrt{\Delta}}{2}\right)^n$ we obtain

$$2 < T' + U'\sqrt{\Delta} \leq T_0 + U_0\sqrt{\Delta},$$

for some integer solution T', U' to the Pell equation, and hence $0 < T' - U'\sqrt{\Delta} < 2$ so (T', U') is a positive solution. As (T_0, U_0) is the minimal such solution, we must have equality everywhere. □

It follows that the stabiliser of F is isomorphic to $\{\pm 1\} \times \mathbf{Z}$, generated by the *automorph* of F:

$$\gamma_F = \begin{pmatrix} \frac{T_0 - bU_0}{2} & aU_0 \\ -cU_0 & \frac{T_0 + bU_0}{2} \end{pmatrix},$$

where (T_0, U_0) is the *fundamental solution* of the Pell equation $T^2 - \Delta U^2 = 4$.

2.4 Parabolic quadratic forms.

This case is usually ignored, most likely as it is not as deep or rich in applications as the previous two cases. Let us nonetheless attempt to formulate a reduction theory ourselves. A parabolic form $F = \langle a, b, c \rangle$ is called reduced if $a = b = 0$. We have

Theorem 2.5. *Any parabolic form is equivalent to a unique reduced form.*

Proof. Clearly, any parabolic form must represent 0, and is therefore equivalent to a form $\langle 0, b, c \rangle$. We have $T^k \cdot \langle 0, b, c \rangle = \langle 0, b, c + kb \rangle$, so if $b \neq 0$ we see that our forms must take both positive and negative values, which is a contradiction, whence $b = 0$. Finally, it is easy to see that $\langle 0, 0, c \rangle \sim \langle 0, 0, c' \rangle$ implies $c = c'$. □

Automorphisms. We showed above that any parabolic form is equivalent to a reduced form, whose automorphisms are just those elements of $\mathrm{SL}_2(\mathbf{Z})$ of the form

$$\gamma = \pm \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix},$$

for any a in $\mathbf{Z}$. This shows that just as in the indefinite case, the stabiliser is isomorphic to $\mathrm{C}_2 \times \mathbf{Z}$.

3 Conway's topograph

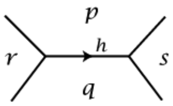
Conway [Con97] presents a convenient visual method for investigating quadratic forms. Not only is it extremely enlightening to think of an equivalence class of quadratic form this way, it often leads to short and clear proofs. All the pictures here are taken from the excellent treatment by Hatcher [Hat17]. I highly recommend reading the first chapter of Conway's book [Con97] for more details and further pictures.

It is well–known that

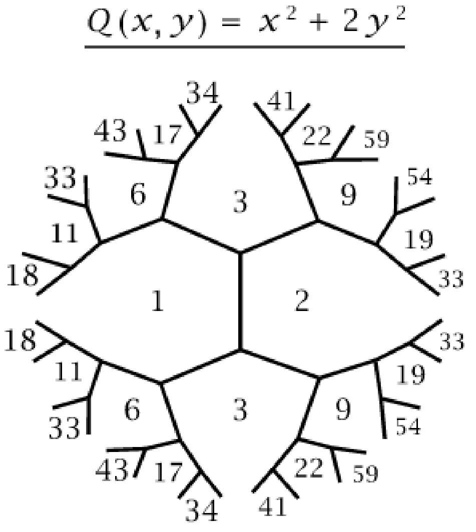
$$\mathrm{SL}_2(\mathbf{Z}) \simeq \mathrm{C}_4 *_{\mathrm{C}_2} \mathrm{C}_6,$$

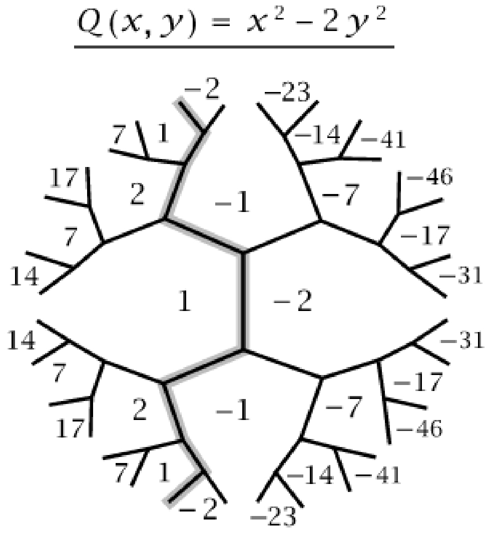
and such an amalgam corresponds to an action of $\mathrm{SL}_2(\mathbf{Z})$ on a tree by Bass–Serre theory [Ser80], which in this case is a 3-regular tree. We may interpret this tree in concrete terms as follows: Its edges correspond to bases of a $\mathbf{Z}$-lattice of rank 2 up to sign, and two edges share a vertex if they share one of their two basis vectors up to sign. We may embed this tree into the plane such that edges that share a basis vector are all adjacent to the same 'region' in the plane. The 3-regular tree embedded in the plane in this manner is called the *topograph.* Given a quadratic form F and a vector in $\mathbf{Z}^2$, we the value of F on this vector to its corresponding region of the plane. Furthermore, to every edge we assign the second coefficient of F in the corresponding basis. We assign an orientation to every edge to keep track of orderings of the basis. The convention is that the value assigned to an edge is always positive, and the region to the left corresponds to the first basis vector.

Starting with $F = \langle p, h, q \rangle$, we construct the topograph step by step as follows. Start with an arbitrary edge, and attach p and q to the two regions bordering the edge, and attaching h to the edge itself, as well as the orientation for which p is on the left, and q is on the right.



Now proceed by computing $s = p + q + h$ and $r = p + q - h$. The value of the edge between p and s is the positive square root of $\Delta + 4ps$, and its orientation is pointing towards q if $q > p + s$ and away from q otherwise. Here are two examples of topographs as shown in Hatcher [Hat17]. Complete this picture by determining the orientations of the edges:

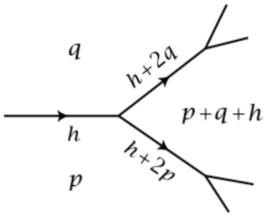




To turn our practical computations into rigorous proofs, the following simple lemma often suffices.

Lemma 3.1 (Climbing lemma). *Suppose q, p, and h in the figure above are positive. Then the number s is also positive, and the edges adjacent to s point away from the vertex shared by p, q, and s.*

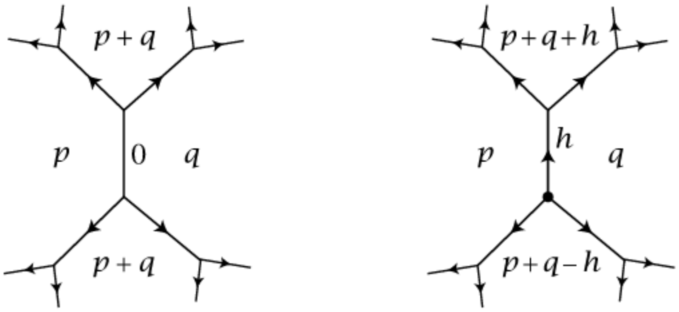
The proof of this lemma is a very easy verification, and it gives us the following schematic rendition of the topograph locally around an edge where all the numbers are positive:



The shape of the topograph reveals much of the inner workings of an orbit of quadratic forms, and can be an extremely enlightening thing to keep in mind. We will now try and be as explicit as possible about the topograph in the various cases considered above.

3.1 Definite forms.

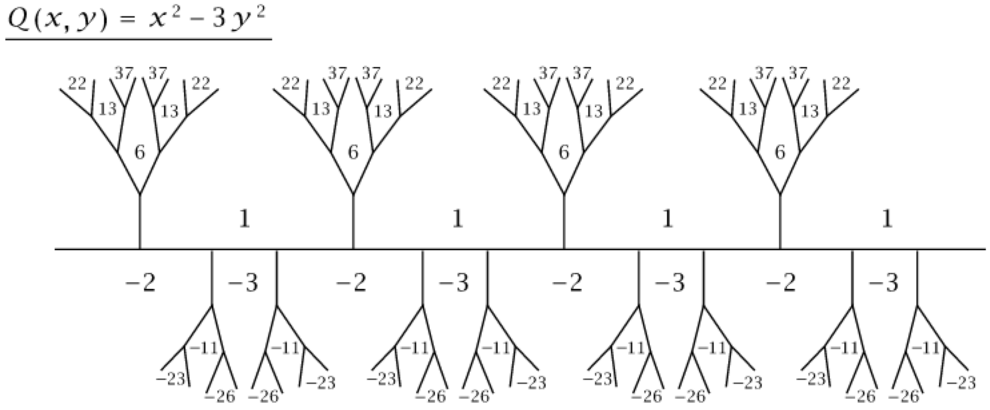
First, we note that there is no essential pictorial difference between postive and negative definite forms, other than that the numbers are all either positive or negative. Therefore, let us assume that we are dealing with a positive definite form. By the climbing lemma, if we follow the flow of the arrows then the values of all regions keep increasing. Retracing our steps, we see that there must be a 'source' or 'well' in the topograph, and the only two possibilities are:



When we discuss Gauß composition below, we will see that if the first case happens, the form must necessarily be of order at most 2 in the class group. But more on that later.

3.2 Indefinite forms.

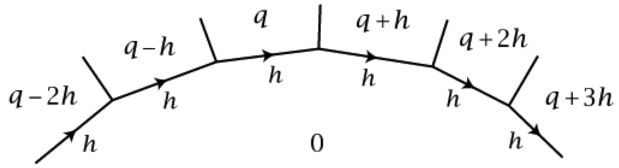
First, assume $\Delta > 0$ is not a square. Then F represents both negative and positive values, but not 0. This implies that there must be an edge adjacent to regions of opposite signs. Following the procedure to complete the topograph starting from this edge, we see that on either side of this edge, there must be another edge with the same property. The edges that separate positive and negative values must therefore be an infinite chain, which Conway calls the *river*.



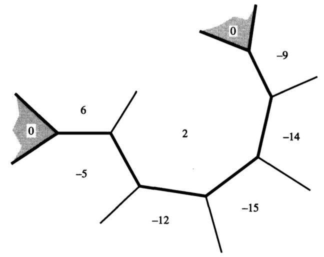
By the climbing lemma, if we move away from the river into the positive side, the values will continuously increase, whereas they will continuously decrease as we venture into the negative side. The river is clearly the most interesting region. Note that the condition $pq < 0$ means that $h^2 - 4pq = \Delta$ only has a finite number of solutions, and hence the river must eventually become periodic! This means that the topograph has non-trivial translation symmetries, corresponding to matrices in $\mathrm{SL}_2(\mathbf{Z})$ that fix our given quadratic form. We now **see** that the stabiliser is projectively infinite cyclic, and hence isomorphic to $\pm\gamma^{\mathbf{Z}}$. The automorph is precisely the transformation that corresponds to translating by the period.

But we are not done yet! Recall our curious algebraic definition of a reduced form in this setting. This is exactly a form corresponding to an edge of the river where the trees hanging off the river switch from the negative side to the positive side, or conversely. In the above picture of the form $x^2 - 3y^2$ we obtain exactly 2 reduced forms.

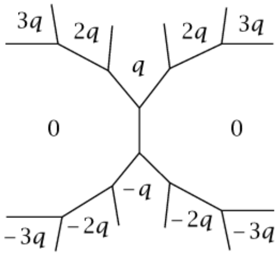
Finally, assume $\Delta = h^2$ is a square, and say h is positive. The quadratic form has rational roots and hence represents 0. There is therefore a region labelled 0, which Conway calls a *lake*:



Because we are in the case where $h \neq 0$, there are two distinct rational roots, and hence there are exactly two lakes in the topograph. As in the above picture, the regions adjacent to the lake have values that form an arithmetic progression, and hence they must change sign at some step. This shows that they must each sprout off a river, which then necessarily connect and form the following picture:

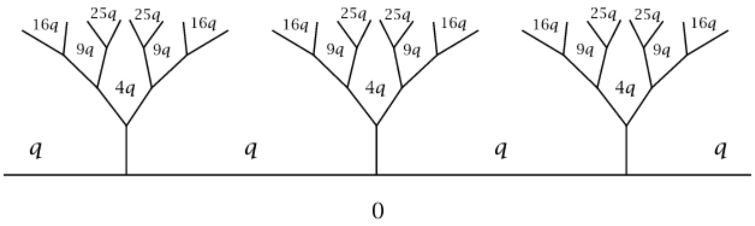


It is possible, as the example $\langle 0,2,0\rangle$ shows, that the length of the river between the two lakes is zero, in which case the two lakes share an edge. See Conway [Con97] for more pictures and details.



3.3 Parabolic forms.

Parabolic forms again represent 0, so they must have a lake. However, the arithmetic progression of values adjacent to the lake cannot change sign and must therefore be constant. This yields the following picture:



4 The generalised Pell equation

For a positive discriminant Δ, we have previously discussed the Pell equation. We now revisit this discussion using the topograph, and give a general procedure for solving the *generalised Pell equation*

$$x^2 - ny^2 = a,$$

where n is some arbitrary integer. We have seen that when $a = 4$, the existence of a solution is always guaranteed. However, for general a this question is very subtle and has no well-understood answer. The only thing we can do is check any given example using the procedures outline here.

Theorem 4.1. *Let $n > 0$ be a non-square in $\mathbf{Z}$, then for any a the equation $x^2 - ny^2 = a$ has at most finitely many equivalence classes of solutions in $\mathbf{Z}$, where we consider two solutions (x, y) and (x', y') equivalent if*

$$x + y\sqrt{n} = (x' + y'\sqrt{n})u, \qquad \textit{for some } \ u \in \mathbf{Z}[\sqrt{n}]^\times \ \textit{ with } \mathrm{Nm}(u) = 1.$$

But how do we find these equivalence classes explicitly, and solve the generalised Pell equation completely? We need an effective way to enumerate equivalence classes of solutions and find at least one solution in every class. This can be done purely algebraically, as in the next theorem.

Theorem 4.2. *Let $n > 0$ be a non-square in $\mathbf{Z}$, then any solution to the equation $x^2 - ny^2 = a$ is equivalent to a solution in the bounded region of $\mathbf{R}^2$ defined by*

$$|x| \leq \sqrt{|a|u}, \ \ |y| \leq \frac{\sqrt{|a|/u}}{\sqrt{n}},$$

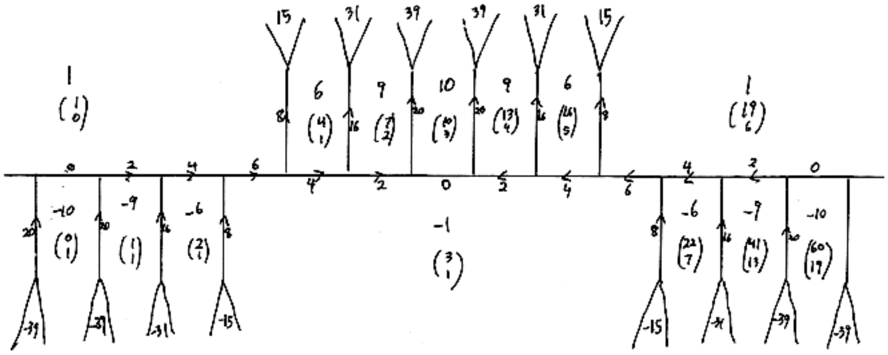
where $u \in \mathbf{Z}[\sqrt{n}]^\times \backslash \{\pm 1\}$ with $\mathrm{Nm}(u) = 1$ and $u > 1$.

These results admit perfectly fine algebraic proofs, though they are quite tedious and perhaps do not give one a true sense for why this must be true. Instead of looking at the proofs, we will interpret the question in terms of Conway's topograph, and see that both results have obvious counterparts in our diagrams which may be proved by considerations with the periodicity of the river and the climbing lemma.

4.1 Using the topograph to solve Pell equations.

We can use the topograph to visualise this method of finding all solutions to certain generalised Pell equations. Perhaps this is best explained on some examples, which together with the pictures below are due to Will Jagy.

Example. Consider the generalised Pell equation $x^2 - 10y^2 = 9$. Start by drawing the "river" of the corresponding Conway topograph:

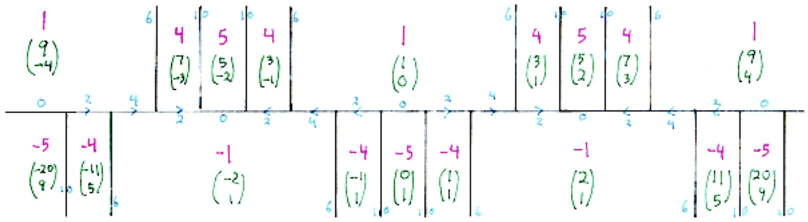


Note that in addition to the usual numbers and orientations, also the basis vector corresponding to every region is drawn in green. As soon as we hit a region with the number 9 in it, we will immediately know the corresponding solution from this basis. We do not need to draw any more than this picture, as the numbers in the trees hanging off the river will keep growing in absolute value by the climbing lemma above. First, we have the non-primitive solutions coming from all solutions to the usual Pell equation $x^2 - 10y^2 = 1$, multiplied by ±3. So now let us consider primitive solutions of $x^2 - 10y^2 = 9$, which may be obtained from $(7, 2)$ and $(13, 4)$ by shifting along the river, which is the same as multiplying by the automorph

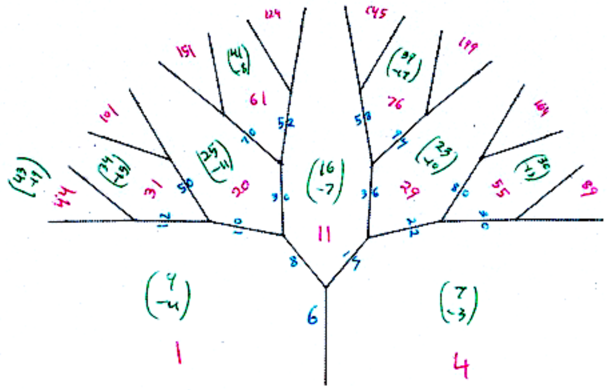
$$\gamma = \begin{pmatrix} 19 & 60 \\ 6 & 19 \end{pmatrix}.$$

In other words, all solutions are $\left\{\pm\gamma^{\mathbf{Z}} \cdot \binom{7}{2}, \pm\gamma^{\mathbf{Z}} \cdot \binom{13}{4}, \pm\gamma^{\mathbf{Z}} \cdot \binom{3}{0}\right\}$.

Example. Consider the generalised Pell equation $x^2 - 5y^2 = 44$. To find all solutions, we first draw the river of the corresponding topograph. We know by the results in the previous section that the river is periodic, and hence completely determined by the following picture.



Note that in addition to the usual numbers and orientations, also the basis vector corresponding to every region is drawn in green. As soon as we hit a region with the number 11 or 44 in it, we will immediately know the corresponding solution from this basis. If we continue to draw the trees in this diagram, we quickly find that there are precisely two reflected trees that contain the numbers 11 and 44, one of which is



We need not consider any more edges by the climbing lemma. The automorph is easily calculated from the above picture to be

$$\gamma = \begin{pmatrix} 9 & 20 \\ 4 & 9 \end{pmatrix}.$$

All imprimitive solutions come from primitive solutions of $x^2 - 5y^2 = 11$, and from the above drawings we deduce that all solutions are given by

$$\left\{ \pm\gamma^{\mathbf{Z}} \cdot \begin{pmatrix} 43 \\ -19 \end{pmatrix}, \pm\gamma^{\mathbf{Z}} \cdot \begin{pmatrix} 43 \\ 19 \end{pmatrix}, \pm\gamma^{\mathbf{Z}} \cdot \begin{pmatrix} 32 \\ 14 \end{pmatrix}, \pm\gamma^{\mathbf{Z}} \cdot \begin{pmatrix} 32 \\ -14 \end{pmatrix} \right\}.$$

More general generalised Pell equations. Now that we have a very clear visual picture of the solutions of these types of equations, it should be clear that essentially the same algorithm allows us to solve

$$F(x, y) = a,$$

for any quadratic form F. Try to experiment with different types of quadratic forms, such as more general indefinite forms, definite forms, parabolic forms etc. See also the exercises for a number of concrete equations.

5 Gauß composition and Bhargava cubes

Now that we have a good understanding of the $\mathrm{SL}_2(\mathbf{Z})$-action on quadratic forms, we introduce a new tool: composition of equivalence classes. For Gauß, the existence of a natural group law on classes was one of the reasons for considering $\mathrm{SL}_2(\mathbf{Z})$-orbits, and not $\mathrm{GL}_2(\mathbf{Z})$-orbits as was more natural from the viewpoint of the question which primes are represented by a certain form.

Gauß defined a quadratic form $H(x, y)$ to be a *direct composition* of quadratic forms $F(x, y)$ and $G(x, y)$ if there exist two bilinear forms

$$\begin{cases} B_1(x,y,z,w) &= a_1xz + b_1xw + c_1yz + d_1yw \\ B_2(x,y,z,w) &= a_2xz + b_2xw + c_2yz + d_2yw, \end{cases}$$

such that $F(x, y)G(z, w) = H(B_1(x, y, z, w), B_2(x, y, z, w))$, and

$$\begin{cases} a_1b_2 - a_2b_1 = F(1,0), \\ a_1c_2 - a_2c_1 = G(1,0). \end{cases}$$

Clearly, this notion formalises the identity that was of crucial importance for the descent step in Euler's original argument! Gauß went on to prove many amazing theorems about the existence of direct compositions and its properties, most notably that it endows the set of equivalence classes of quadratic forms with the structure of a finite abelian group which is called the *class group*. This composition is perhaps most familiar to you in the language of ideals, which we will adopt soon, and since the proofs provided by Gauß are complicated and difficult, it is often overlooked in the setting of quadratic forms. The following result is clear from the definition:

Lemma 5.1. *Suppose Q_1 and Q_2 are quadratic forms that represent m and n respectively, and suppose Q_3 is a direct composition of Q_1 and Q_2. Then Q_3 represents mn.*

Clearly, knowing that some form is a direct composition of two other forms is a useful piece of information when talking about representability. However, from the definition it is not clear that we may find a direct composition of any two forms, or whether any two such compositions are always $\mathrm{SL}_2(\mathbf{Z})$-equivalent.

5.1 Bhargava cubes.

In 2004, Bhargava presented a new treatment of Gauß composition that clears up Gauß' presentation, and, more importantly, allows for generalisations to new settings. We define a Bhargava cube to be a $2 \times 2 \times 2$ cube with integers associated to its vertices. To a Bhargava cube, we associate three

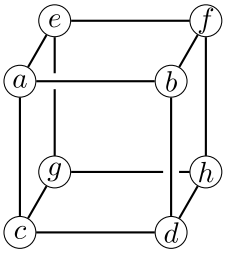


Figure I.1: Cubus Bhargaviensis

quadratic forms as follows:

$$\begin{cases} Q_1(x,y) &= -\det\left(x\cdot\begin{pmatrix} a & e\\ b & f\end{pmatrix} + y\cdot\begin{pmatrix} c & g\\ d & h\end{pmatrix}\right)\\ Q_2(x,y) &= -\det\left(x\cdot\begin{pmatrix} a & c\\ e & g\end{pmatrix} + y\cdot\begin{pmatrix} b & d\\ f & h\end{pmatrix}\right)\\ Q_3(x,y) &= -\det\left(x\cdot\begin{pmatrix} a & b\\ c & d\end{pmatrix} + y\cdot\begin{pmatrix} e & f\\ g & h\end{pmatrix}\right)\end{cases}$$

These three forms have the same discriminant. If two of these three quadratic forms are primitive, then so is the third one. In this case, we say the cube is *projective*. It turns out that $Q_3(x,-y)$ is a direct composition of $Q_1(x,y)$ and $Q_2(x,y)$ in the sense of Gauß! The language of Bhargava has the advantage of "unraveling" some of the difficult algebra of Gauß. Of course the main virtue of Bhargava cubes is that they suggest generalisations which genuinely go beyond the work of Gauß. All results we discuss here, and even to some extent the idea of using cubes, are due to Gauß, Dirichlet, and others, but we find Bhargava's language too elegant and convenient to avoid.

Symmetries of a cube. The symmetry group of a cube has order 48. For every symmetry, we wonder what its effect is on the three quadratic forms. Suppose we have a Bhargava cube $\mathcal{A}$ which gives rise to the quadratic forms $Q_i = \langle A_i, B_i, C_i\rangle$.

- Rotation by $\frac{2\pi}{3}$ around ah: This changes (Q_1, Q_2, Q_3) to (Q_3, Q_1, Q_2).
- Rotation by $\frac{\pi}{2}$ around vertical axis: This changes (Q_1, Q_2, Q_3) to (Q_1', Q_2', Q_3'), where

$$\begin{cases} Q_1' = \langle -A_1, -B_1, -C_1\rangle,\\ Q_2' = \langle C_3, B_3, A_3\rangle,\\ Q_3' = \langle -A_2, -B_2, -C_2\rangle\end{cases}$$

- Reflection switching front and back: This changes (Q_1, Q_2, Q_3) to (Q_1', Q_2', Q_3'), where

$$\begin{cases} Q_1' = \langle -A_1, -B_1, -C_1\rangle,\\ Q_2' = \langle -A_2, -B_2, -C_2\rangle,\\ Q_3' = \langle C_3, B_3, A_3\rangle\end{cases}$$

Action of $\mathrm{SL}_2(\mathbf{Z})\times\mathrm{SL}_2(\mathbf{Z})\times\mathrm{SL}_2(\mathbf{Z})$ on Bhargava cubes. We see that we may realise the action of the modular group on quadratic forms as various actions on Bhargava cubes. As any Bhargava cube gives rise to three quadratic forms, there are several ways to do this, yielding an action of the triple product on cubes. More precisely, let $\mathcal{A}$ be a Bhargava cube. A matrix

$$\gamma = \begin{pmatrix} q & r\\ s & t\end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$$

acts on $\mathcal{A}$ by replacing the cube $\mathcal{A}$ with top face M_1 and bottom face N_1 by the cube with top face $qM_1 + sN_1$ and bottom face $rM_1 + tN_1$. On the three attached quadratic forms, this induces the usual action of $\mathrm{SL}_2(\mathbf{Z})$ on Q_1, and acts trivially on Q_2 and Q_3. Using similar actions on the left/right and front/back faces, we get an action of the group

$$\mathrm{SL}_2(\mathbf{Z}) \times \mathrm{SL}_2(\mathbf{Z}) \times \mathrm{SL}_2(\mathbf{Z})$$

on the set of Bhargava cubes, which preserves the set of projective Bhargava cubes.

Group laws. We say that three primitive quadratic forms Q_1, Q_2, Q_3 are *collinear* if there is a projective Bhargava cube $\mathcal{A}$ with precisely those three associated forms. We now come to the main result of Gauß composition, which shows us that the set of equivalence classes of quadratic forms naturally carries a group structure.

Theorem 5.1 (Gauß). *Given two primitive forms Q_1, Q_2 with the same discriminant Δ, there is always a form Q_3 with discriminant Δ such that Q_1, Q_2, Q_3 are collinear, and any two such forms are equivalent. Declaring collinear triples to have product 1 makes the set of equivalence classes of primitive forms of discriminant Δ into a finite abelian group.*

We will not prove this theorem, but note that the proof uses elementary but ingenious congruences and is completely constructive and explicit. Ignoring the fact that the language of group theory was not used at the time, all proofs were given by Gauß in his *Disquisitiones Arithmeticae.* The trivial class for Gauß composition is that of the *principal form* $\langle 1, \sigma, m\rangle$. We may find the inverse class of $\langle a, b, c\rangle$ by considering the cubes

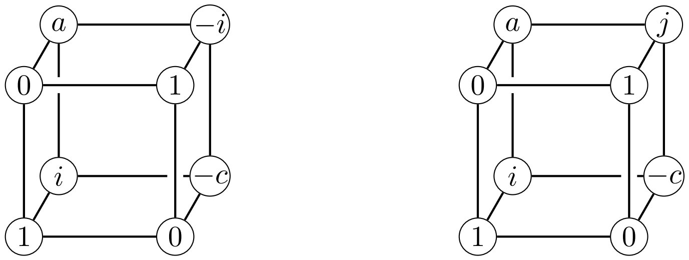


Figure I.2: The inverse class

where $\Delta = -4m + \sigma$ with $\sigma \in \{0, 1\}$, $2i = b + \sigma$ and $j = 1 - i$. Here, we consider the left cube if $\sigma = 0$ and the right cube if $\sigma = 1$. We calculate that the third quadratic form attached to this cube is the principal form, whereas the other two are $\langle a, b, c\rangle$ and $\langle a, -b, c\rangle$. This shows that the inverse class of any quadratic form $\langle a, b, c\rangle$ is that of $\langle a, -b, c\rangle$. It is also possible to define a group law on the set of projective Bhargava cubes, but we will say nothing about this here.

5.2 Narrow and wide class groups.

As we have just seen, the equivalence classes $\mathrm{Cl}^+(\Delta)$ of quadratic forms of discriminant Δ carry a natural group structure. Later, we will make the connection with class groups of orders in quadratic number fields, so let us make a few modification in preparation of that.

We first define an action of $\mathrm{GL}_2(\mathbf{Z})$ on quadratic forms which is **different** from the action considered before. The notion above is natural from the point of view of representations of integers by quadratic forms, and is the one considered by most pre-Gauß mathematicians. If we want to obtain a group structure on such improper equivalence classes however, we need to modify the action by setting

$$\gamma = \begin{pmatrix} q & r \\ s & t \end{pmatrix} \in \mathrm{GL}_2(\mathbf{Z}), \qquad \text{then } \gamma \cdot F(x,y) = \frac{1}{\det\gamma} \cdot F(qx+ry, sx+ty).$$

From now on, whenever we consider an action of $\mathrm{GL}_2(\mathbf{Z})$ on quadratic forms, it will be this one. The $\mathrm{GL}_2(\mathbf{Z})$-orbits of primitive forms of discriminant Δ are denoted by $\mathrm{Cl}(\Delta)$. There is a natural map

$$\pi : \mathrm{Cl}^+(\Delta) \longrightarrow \mathrm{Cl}(\Delta)$$

that sends any $\mathrm{SL}_2(\mathbf{Z})$-class to its $\mathrm{GL}_2(\mathbf{Z})$-orbit. As the former group is of index 2 in the latter, the fibres of this map have order at most 2. We can use this map to give a group structure to $\mathrm{Cl}(\Delta)$, as follows. Take representatives Q_1 and Q_2 of two classes in $\mathrm{Cl}(\Delta)$, and define the product of these two classes to be $\pi(Q_1 \cdot Q_2)$, where the product is taken in $\mathrm{Cl}^+(\Delta)$. To check that this is well-defined, we only need to check that the product of $Q_1 = \langle a,b,c\rangle$ and $Q_2 = \langle d,e,f\rangle$ is $\mathrm{GL}_2(\mathbf{Z})$-equivalent to the product of $Q_1^* = \langle -a,b,-c\rangle$ and $Q_2^* = \langle -d,e,-f\rangle$. To show this, take a Bhargava cube $\mathcal{A}$ whose first and second quadratic forms are Q_1 and Q_2, and whose third form is $Q_3 = \langle t,u,v\rangle$, a composite of the two. Now swap the front and back of this cube, and take inverses of all the forms that we obtain. This shows that Q_1^*, Q_2^*, and Q_3 are collinear. See also one of the exercises in this chapter. Said exercise furthermore proves that there is an exact sequence

$$1 \to \{\pm 1\}/\mathrm{Nm} \longrightarrow \mathrm{Cl}^+(\Delta) \longrightarrow \mathrm{Cl}(\Delta) \to 1.$$

Here, the group Nm is equal to $\{\pm 1\}$ if the principal form represents -1, and trivial otherwise. This sequence is not always split, see the exercises.

6 Genus theory and Quadratic Reciprocity

Now that we have studied quadratic forms up to $\mathrm{SL}_2(\mathbf{Z})$-equivalence, we return to our initial question of how to generalise Fermat's theorem on primes of the form $x^2 + y^2$. In general, can we give a clean criterion for when a prime p is represented by an arbitrary quadratic form F? We start with the weaker question:

> Which values in $(\mathbf{Z}/\Delta\mathbf{Z})^\times$ are represented by a given form of discriminant Δ?

Genus theory will provide a full answer to this question, which often suffices to answer the stronger question of primes represented by a given form, though as we will see, it is not in general sophisticated enough. A great reference for the material in this section is Flath [Fla88, Chapter 5].

6.1 Elementary genus theory.

We begin with some elementary consequences of the theory of Gauss composition, motivated by the fact that if two quadratic forms Q_1, Q_2 of the same discriminant represent m, n respectively, then their composite represents mn. This gives a group-like structure on the set of values represented by quadratic forms of a certain discriminant, which becomes an honest group structure upon reducing modulo Δ. More precisely, we have the following result:

Lemma 6.1. *All elements in $(\mathbf{Z}/\Delta\mathbf{Z})^\times$ that are represented by the principal form, form a subgroup H. Let F be any quadratic form of discriminant Δ, then all elements in $(\mathbf{Z}/\Delta\mathbf{Z})^\times$ that are represented by F form a coset of H.*

Proof. This follows immediately from Gauß composition. For the first part, observe that the principal form always represents 1, and Gauß composition shows that H is closed under products. Furthermore, if $a \in (\mathbf{Z}/\Delta\mathbf{Z})^\times$ is represented by the principal form, then a^n is as well, proving that H contains inverses as every element is of finite order. For the second part, Gauß composition shows us that if $a \in (\mathbf{Z}/\Delta\mathbf{Z})^\times$ is represented by F, and t is the order of F in the class group, then $a^t = h \in H$ and the element $a^{-1}h$ is represented by F^{-1}. It follows that if a and b are represented by F, then $a^{-1}b \in H$. Furthermore, any quadratic form represents values coprime to Δ, so the set of values represented by F is non-empty and hence a coset of H. □

We say that two primitive quadratic forms of discriminant Δ are in the same *genus* if they represent the same values in $(\mathbf{Z}/\Delta\mathbf{Z})^\times$. Clearly, equivalent forms are in the same genus, but in general genera can consist of several classes. We obtain a map

$$\Phi : \mathrm{Cl}^+(\Delta) \to (\mathbf{Z}/\Delta\mathbf{Z})^\times/H$$

that sends every class to the coset of H represented by any representative of the class. This is clearly a group homomorphism of abelian groups, whose target is of a very simple nature, as we will see below.

6.2 A proof of quadratic reciprocity.

Many proofs of quadratic reciprocity are known today, but in this context one of them stands out. Amazingly, Gauß used the concept of genera to give a proof of quadratic reciprocity, which we will now sketch. Of the four proofs he gave in his *Disquisitiones Arithmeticae*, this is arguably the deepest one. It proceeds via the following steps

- Step 1: Show that there are at most $2^{\mu-1}$ genera in discriminant Δ, where μ is defined below.
- Step 2: Deduce quadratic reciprocity by considering forms of discriminant p, q, and pq.

Let r be the number of odd primes dividing Δ. We define the number μ by

$$\mu = \begin{cases} r & \text{if} \quad \Delta \equiv 1, 4, 5, 9, 13, 17, 20, 21, 25, 29 \pmod{32} \\ r+1 & \text{if} \quad \Delta \equiv 8, 12, 16, 24, 28 \pmod{32} \\ r+2 & \text{if} \quad \Delta \equiv 0 \pmod{32}. \end{cases}$$

We will write down a set of μ characters on $(\mathbf{Z}/\Delta\mathbf{Z})^\times$ such that the intersection of their kernels is precisely H. More precisely, define the characters

$$\chi_i := \left(\frac{\cdot}{p_i}\right), \quad \delta := (-1)^{(\cdot - 1)/2}, \quad \epsilon := (-1)^{(\cdot^2 - 1)/8},$$

where $p_1, p_2, \ldots, p_r$ are the odd primes dividing Δ. Taking the first r of these, as well as a combination of δ and ϵ depending on the residue of Δ modulo 32, we obtain a homomorphism

$$\Psi : (\mathbf{Z}/\Delta\mathbf{Z})^\times \longrightarrow \{\pm 1\}^\mu$$

whose kernel we can check in all cases to be equal to H. We refer the reader to Flath [Fla88] for details. The argument breaks down into a large number of cases, where most of the difficulty lies in dealing with powers of 2. As far as I know, no universal argument exists for all cases, but let me know if you do find one. We will discuss only the case where Δ is odd, in which case the r characters we consider are

$$\chi_i : \left(\frac{\cdot}{p_i}\right), \quad 1 \le i \le r.$$

The elements that map to 1 under all these characters are precisely the squares in $(\mathbf{Z}/\Delta\mathbf{Z})^\times$, since by the Chinese remainder theorem the latter group is the product of the groups $(\mathbf{Z}/p_i^{n_i}\mathbf{Z})^\times$ which are cyclic of order $p_i^{n_i - 1}(p-1)$ so that its squares are precisely the elements that map to a square in $(\mathbf{Z}/p_i\mathbf{Z})^\times$. On the other hand, the set of squares in $(\mathbf{Z}/\Delta\mathbf{Z})^\times$ is precisely the set of elements represented by the principal form, as

$$4\left(x^2 + xy + \frac{1-\Delta}{4}y^2\right) \equiv (2x+y)^2 \pmod{\Delta}.$$

This proves the claim that the r characters under consideration cut out the group H in the case where Δ is odd. When Δ is even, the analysis is more involved because of the presence of the prime 2, but similar arguments using slightly different sets of characters work in those cases as well.

This tells us that H has index at most 2^μ in $(\mathbf{Z}/\Delta\mathbf{Z})^\times$. Now consider the aformentioned map $\Phi : \mathrm{Cl}^+(\Delta) \longrightarrow (\mathbf{Z}/\Delta\mathbf{Z})^\times/H$ which sends a class to the coset of H it represents. As this map is not surjective, it follows that there are at most $2^{\mu-1}$ genera of discriminant Δ.

Now, let p, q be distinct odd prime numbers, and set $p^* = (-1)^{(p-1)/2}p$ and $q^* = (-1)^{(q-1)/2}q$. There are two cases to consider for proving quadratic reciprocity:

- The case $\left(\frac{p^*}{q}\right) = 1$. This implies that q is represented by some form of discriminant p^*, which is odd and hence there is a unique genus. This then implies that q needs to be in the kernel of the associated character, which is to say that

$$\left(\frac{q}{p}\right) = 1.$$

- The case $\left(\frac{p^*}{q}\right) = -1$. Suppose first that p or q is 1 modulo 4, then $\left(\frac{p^*}{q}\right) = \left(\frac{p}{q}\right)$ and so by the previous case we must have $\left(\frac{q^*}{p}\right) = -1$. Suppose that $p \equiv q \equiv 3 \pmod 4$, then set $\Delta = pq$ and consider the map

$$\Psi : (\mathbf{Z}/\Delta\mathbf{Z})^\times \longrightarrow \{\pm 1\}^\mu.$$

As Δ is odd, we know $\mu = 2$ and the map Ψ is constructed using the two associated characters modulo p and q. In particular, there are at most 2 genera. Consider the forms

$$\langle p, p, \frac{p-q}{4}\rangle, \quad \langle -p, -p, \frac{q-p}{4}\rangle$$

which are of discriminant pq. The first represents $p-q$, and the second represents $q-p$, so that the cosets of H they represent in $(\mathbf{Z}/\Delta\mathbf{Z})^\times$ both have different images under Ψ, and hence at least one of them is in the principal genus. The second form cannot be in the principal genus by the assumption $\left(\frac{p^*}{q}\right) = \left(\frac{-p}{q}\right) = -1$, and so the former must be in the principal genus, which implies that

$$\left(\frac{p-q}{p}\right) = \left(\frac{-q}{p}\right) = 1.$$

6.3 More genus theory.

First, we recall the *Jacobi symbol*, which is a generalisation of the familiar Legendre symbol. Given $n > 0$ relatively prime to a, we define

$$\left(\frac{a}{n}\right) = \prod_{i=1}^{r} \left(\frac{a}{p_i}\right),$$

where $n = p_1 \cdots p_r$ is the prime factorisation of n. This symbol has good multiplicative properties, concisely packaged into the following statement which is of fundamental importance in genus theory:

Theorem 6.1. *There is a unique homomorphism*

$$\chi : (\mathbf{Z}/\Delta\mathbf{Z})^\times \longrightarrow \{\pm 1\},$$

such that for any prime $p \nmid 2\Delta$ we have

$$\chi(p) = \left(\frac{\Delta}{p}\right), \text{ and } \chi(-1) = \begin{cases} 1 & \text{if } \Delta > 0 \\ -1 & \text{if } \Delta < 0 \end{cases}.$$

Any prime $p \nmid 2\Delta$ is represented by some quadratic form of discriminant Δ if and only if $p \in \operatorname{Ker} \chi$.

Proof. The existence of χ can be deduced by repeated applications of quadratic reciprocity. In fact, the existence of χ is equivalent to the statement of quadratic reciprocity. For the second statement, note that any form F that properly represents p is equivalent to

$$F \sim \langle p, b, c\rangle,$$

for some b, c as can be checked using an elementary calculation. Therefore, if p is represented by some form with discriminant Δ, then $\Delta = b^2 - 4pc \equiv b^2 \pmod{p}$ and hence $\left(\frac{\Delta}{p}\right) = 1$. Conversely, if $\Delta \equiv b^2 \pmod{p}$ for some b, we may assume that $b \equiv \Delta \pmod{2}$, which implies $\Delta - b^2 = -4c$ for some c. The form $\langle p, b, c\rangle$ is then of discriminant Δ and represents p. □

This shows that in fact the morphism Ψ defined above factors through a surjective morphism

$$\Phi : \mathrm{Cl}^+(\Delta) \to \mathrm{Ker}\ \chi/H$$

The main theorem of genus theory identifies the kernel of this homomorphism:

Theorem 6.2. *The principal genus consists of the classes in* $\mathrm{Cl}^+(\Delta)^2$*, and the number of genera is* $2^{\mu-1}$.

Proof. It is an important fact that the number of *ambiguous classes* in $\mathrm{Cl}^+(\Delta)$, i.e. the elements of order at most 2, is exactly equal to $2^{\mu-1}$. The idea is to generalise exercise 14 and show that any ambiguous form is equivalent to one of the form $\langle a, 0, c\rangle$ or $\langle a, a, c\rangle$, and counting those is a straightforward matter. Assuming this result, the theorem is a consequence of the results in the previous section. Indeed, we see that all squares are represented by the principal form, and hence Φ factors through $\mathrm{Cl}^+(\Delta)/\mathrm{Cl}^+(\Delta)^2$, which has size equal to the number of ambiguous classes, which is $2^{\mu-1}$. On the other hand, there are $2^{\mu-1}$ genera by the above argument, which means that the kernel of Φ must be precisely $\mathrm{Cl}^+(\Delta)^2$. □

6.4 When is a prime of the form $x^2 + ny^2$?

Let us now investigate Euler's conjectures once more. We will do three examples, the first two of which Euler was able to do using the descent and reciprocity steps discussed in the first lecture. The third example was not proved by Euler, and remained a conjecture until the development of genus theory.

- When is $p \neq 2$ of the form $x^2 + 2y^2$? This is a quadratic form of discriminant -8, and one easily checks that $\mathrm{Cl}(-8) = 1$ with reduced form $x^2 + 2y^2$. A prime is represented by some form of discriminant -8, and hence by $x^2 + 2y^2$, if and only if $\left(\frac{-8}{p}\right) = 1$, which translates into Euler's criterion.
- When is $p \neq 3$ of the form $x^2 + 3y^2$? Again, the discriminant is -12 which has wide class number 1. This means that a prime p is represented by the unique class of forms of this discriminant if and only if -12 is a square modulo p. By exercise 1b), this is equivalent to $p \equiv 1 \pmod{3}$.

- Consider the quadratic form $x^2 + 5y^2$, which is of discriminant -20. We calculate that $\mathrm{Cl}(-20) \simeq \mathrm{C}_2$ with reduced forms $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$. Both classes belong to different genera by genus theory, and one quickly sees that a prime p is represented by the principal form $x^2 + 5y^2$ if and only if $p \equiv 1, 9 \pmod{20}$, whereas p is represented by $2x^2 + 2xy + 3y^2$ if and only if $p \equiv 3, 7 \pmod{20}$.

In fact, we can see that this procedure will give us a full solution to the question of when a prime is of the form $x^2 + ny^2$ if and only if the principal genus consists precisely of the principal class. Said differently, genus theory answers this question whenever $\mathrm{Cl}(\Delta)$ is an elementary abelian 2-group.

Now notice two things. Firstly, although we were able to show Euler's conjecture on when p is represented by *either* $x^2 + 14y^2$ *or* $2x^2 + 7y^2$, it should be clear that this is not the end of the story, as we were unable to separate those two forms by congruences, which means that they lie in the same genus, but are not equivalent. Secondly, we do not even come close to an indication of how to prove the conjecture about $x^2 + 27y^2$. The class number is 3, and as is the case with any odd class number greater than one, genus theory is a particularly lousy tool here. More seriously, we have as yet no indication where the mysterious polynomial $t^3 - 2$ comes from. To go beyond genus theory, we will make a significant change of pace, and use class field theory. I will assume everyone has some previous exposure to this topic, and its results will be freely used in what follows. A short outline of the theory, with a strong bias towards concrete applications, may be found in Appendix C.

7 Quadratic orders and ring class fields

We now discuss the connection between quadratic forms and ideals in orders of quadratic extensions of $\mathbf{Q}$. This is the language that is most commonly used to dispense with a systematic discussion of quadratic forms early on. The theory of Gauss composition is in this way avoided in many texts, in favour of the obvious group structure on ideal classes. Both approaches of course have their advantages, and we hope that some insight can be gained from discussing both. The point of view of ideals is the one that will dominate the rest of these notes, and in this section we show how it allows us to invoke class field theory in order to solve our question of when a prime is of the form $x^2 + ny^2$, at least in theory.

7.1 Ideals in quadratic orders.

Let $K = \mathbf{Q}(\sqrt{d})$ be a quadratic number field with d squarefree. Set $\omega = \sqrt{d}$, unless $4 \mid d - 1$ in which case we set $\omega = \frac{1+\sqrt{d}}{2}$. We know that $\mathcal{O}_K = \mathbf{Z} + \mathbf{Z}\omega$. The *order* of conductor f is the unique subring $\mathcal{O}$ of $\mathcal{O}_K$ of index f, which is necessarily equal to $\mathbf{Z} + \mathbf{Z}f\omega$. It has discriminant equal to $f^2\Delta_K$, where Δ_K is the discriminant of the quadratic field K.

Ideals. We want to consider groups of ideals, but there is one problem: General orders are not Dedekind domains, and hence we cannot use the usual properties that ideals enjoy. For instance, we generally do not have unique factorisation into ideals, and fractional ideals are not necessarily invertible for multiplication. To circumvent these issues, we will restrict to a particularly nice class of ideals. Say an ideal $\mathfrak{a} \trianglelefteq \mathcal{O}$ is *proper* whenever

$$\mathcal{O} = \{\alpha \in K \ : \ \alpha\mathfrak{a} \subseteq \mathfrak{a}\}.$$

A *fractional ideal* of $\mathcal{O}$ is a subset of K which is a non-zero finitely generated $\mathcal{O}$-module. Equivalently, it is a subset of the form $\alpha\mathfrak{a}$ where $\alpha \in K^{\times}$ and $\mathfrak{a} \trianglelefteq \mathcal{O}$. We say a fractional ideal is proper if $\mathfrak{a}$ is proper in the above sense, and we say it is *principal* if we may take $\mathfrak{a} = \mathcal{O}$. A fractional $\mathcal{O}$-ideal $\mathfrak{m}$ is said to be *invertible* if there is another fractional $\mathcal{O}$-ideal $\mathfrak{n}$ such that $\mathfrak{m} \cdot \mathfrak{n} = \mathcal{O}$, so that in particular every principal fractional ideal is invertible. It is easy to see that being proper is equivalent to being invertible for fractional ideals. The *narrow ideal class group* of $\mathcal{O}$ is the group formed by proper fractional $\mathcal{O}$-ideals modulo principal ideals generated by an element of K that is positive for every real embedding of K.

Quadratic forms. There is a close relation between the theory of quadratic forms, and the theory of ideals in orders of quadratic $\mathbf{Q}$-algebras. The non-degenerate cases correspond to quadratic number fields, and are the most interesting and difficult instances to consider. More precisely, we may attach to any primitive quadratic form $\langle a, b, c\rangle$ of discriminant Δ an ideal in the order $\mathcal{O}$ of discriminant Δ via

$$\langle a, b, c\rangle \mapsto (a, \frac{-b+\sqrt{\Delta}}{2}) \trianglelefteq \mathcal{O}.$$

Conversely, to an ideal $\mathfrak{a} \trianglelefteq \mathcal{O}$ generated by α, β we attach

$$\frac{\mathrm{Nm}(\alpha x + \beta y)}{\mathrm{Nm}(\mathfrak{a})} = ax^2 + bxy + cy^2.$$

These maps define isomorphisms between $\mathrm{Cl}^{+}(\Delta)$ and the narrow ideal class group of $\mathcal{O}$, and many of the properties of quadratic forms discussed above have corresponding interpretations on the side of ideals, some of which are summarised in the following table:

Quadratic Forms	Ideals
$\langle a, b, c\rangle$	$\mathfrak{a} = \left(a, \frac{b+\sqrt{\Delta}}{2}\right)$
Δ	Discriminant of $\mathcal{O}$
$\mathrm{Cl}^{+}(\Delta)$	Narrow ideal class group
$\mathrm{Cl}(\Delta)$	Ideal class group
Ambiguous forms	$\mathfrak{a} = \sigma(\mathfrak{a})$
$\rho(a, b, c)$ right neighbour	$\frac{-b+\sqrt{\Delta}}{2} \cdot \mathfrak{a}$

Relation to the maximal order. Given that we have translated many of the structures for quadratic forms to the world of proper ideals in certain orders, we are left wondering how one guarantees that a given ideal is in fact proper. One easy class of such ideals is that of those ideals $\mathfrak{a} \trianglelefteq \mathcal{O}$ which are coprime to the conductor, which means that $\mathfrak{a} + f\mathcal{O} = \mathcal{O}$. In fact, those ideals are in some sense enough to understand the entire ideal class group, and furthermore they give us a natural isomorphism between the ideal class group of $\mathcal{O}$ and an ideal class group of ideals in $\mathcal{O}_K$ coprime to f. This is not surprising, as we have already seen a similar situation in the context of the Kummer–Dedekind theorem. In that theorem, we choose an order $\mathbf{Z}[\alpha] \simeq \mathbf{Z}[x]/(f_\alpha)$ in the ring of integers of a number field, in which splitting primes is a straightforward manner because of the explicit presentation of the order. The theorem of Kummer–Dedekind tells us that the splitting of an ideal (p) is not affected by the cornorm map of ideals in the order, to ideals in the full ring of integers, as long as p does not divide the index of the order. For essentially the same reasons, we obtain the following result about ideal class groups:

Lemma 7.1. *Given an order $\mathcal{O}$ of conductor f in $\mathcal{O}_K$, the maps*

$$\mathfrak{a} \mapsto \mathfrak{a}\mathcal{O}_K, \qquad \mathfrak{a} \mapsto \mathfrak{a} \cap \mathcal{O}$$

between fractional $\mathcal{O}$-ideals coprime to f and fractional $\mathcal{O}_K$-ideals coprime to f are mutually inverse and induce an isomorphism between the narrow ideal class group of $\mathcal{O}$ and the class group $I_{(f)}/P_{(f),\mathbf{Z}}$, where $P_{(f),\mathbf{Z}}$ is the set of all principal ideals of $P_{(f)}$ which are congruent to an integer modulo f. More precisely, we have a short exact sequence

$$1 \to (\mathcal{O}_K/f)^\times/\mathcal{O}_K^\times(\mathbf{Z}/f\mathbf{Z})^\times \to \mathrm{Cl}(\mathcal{O}) \to \mathrm{Cl}(\mathcal{O}_K) \to 1.$$

We learn from the above that the class group of an order $\mathcal{O}$ may be interpreted as an explicit quotient of the ray class group $\mathrm{Cl}_{(f)}$, and is therefore naturally the Galois group of some finite abelian extension of K, which we call the *ring class field* attached to $\mathcal{O}$. This is what will allow us to give a class field theoretic description of all the primes that are of the form $x^2 + ny^2$, as we will now see.

7.2 When is a prime of the form $x^2 + ny^2$?

Now that we have introduced, albeit without proof, the main results of class field theory, we can answer our motivating question on when a prime is of the form $x^2 + ny^2$, or any other quadratic form, for that matter. We will formulate a precise theoretical criterion now, though as we will see, it involves one mysterious quantity: The minimal polynomial of a primitive element in a certain ring class field. This is the part of the theory that we will make explicit in the case $\Delta < 0$ using the theory of complex multiplication, in the next chapter.

Theorem 7.1. *Let $n \geq 1$, then there exists a monic irreducible $f_n \in \mathbf{Z}[x]$ of degree $h(-4n)$ such that*

$$p = x^2 + ny^2 \iff \begin{cases} \left(\frac{-n}{p}\right) = 1 & \text{and} \\ f_n \text{ has a root in } \mathbf{F}_p. \end{cases}$$

whenever $(p, 2n \cdot \Delta_{f_n}) = 1$.

Proof. Consider the order $\mathcal{O} = \mathbf{Z}[\sqrt{-n}]$ in $K = \mathbf{Q}(\sqrt{-n})$, which has discriminant $-4n$. If p does not divide $2n$, then the statement that p be of the form $x^2 + ny^2$ is equivalent to the statement that the proper ideal $p\mathcal{O}$ is the product of two principal prime ideals in $\mathcal{O}$. As $p\mathcal{O}$ is proper, we have by the results above that this is equivalent to the statement that $p\mathcal{O}_K = \mathfrak{p}\overline{\mathfrak{p}}$, where $\mathfrak{p}$ is a principal prime ideal congruent to an integer modulo f, meaning its image in $I_{(f)}/P_{(f),\mathbf{Z}}$ is trivial. By class field theory, this is equivalent to the statement that p splits completely in the ring class field L/K of conductor f.

The ring class field L is contained in $K_{(f)}$ by construction, which is Galois over $\mathbf{Q}$ as (f) is stable under $\mathrm{Gal}(K/\mathbf{Q})$, which is generated by complex conjugation τ. From the properties of the Artin map in the appendix, we see that $K_{(f)}$ has Galois group over $\mathbf{Q}$ equal to the semi-direct product $\mathrm{Cl}_{(f)} \rtimes \langle\tau\rangle$, where τ acts by inversion on the abelian group $\mathrm{Cl}_{(f)}$. This leaves the subgroup $(\mathbf{Z}/f\mathbf{Z})^\times \mathcal{O}_K^\times$ of $I_{(f)}/P_{(f)}$ invariant, so that $L/\mathbf{Q}$ is Galois with group $\mathrm{Cl}(\mathcal{O}) \rtimes \langle\tau\rangle$ acting by inversion. This implies that the field $L \cap \mathbf{R}$ is a finite extension of degree $h(-4n)$.

Choose a integral generator α of $L \cap \mathbf{R}$, with minimal polynomial $f_n(x) \in \mathbf{Z}[x]$. The prime p being split in K is equivalent to $\left(\frac{-n}{p}\right) = 1$, whereas p splits completely in $L \cap \mathbf{R}$ if and only if f_n has a root modulo p, whenever p does not divide the discriminant of f_n, by Kummer–Dedekind. □

Example. Consider once more the following conjecture made by Euler:

$$p = x^2 + 27y^2 \iff \begin{cases} p \equiv 1 \pmod 3 \text{ and} \\ t^3 - 2 \in \mathbf{F}_p[t] \text{ has a root.} \end{cases}$$

Can we prove this theorem? By Theorem 7.1, we would have a criterion for when p is of the form $x^2 + 27y^2$ as soon as we find a generator for the ring class field of the order $\mathbf{Z}[\sqrt{-27}]$, which is of conductor 6 in $\mathcal{O}_K$, where $K = \mathbf{Q}(\sqrt{-3})$. First, we calculate easily that there are precisely three reduced forms of discriminant $-4 \cdot 27$, so that the ring class field F/K is a degree 3 extension unramified outside 6. As $\zeta_3 \in K$, we conclude by Kummer theory that

$$F = K(\sqrt[3]{a}),$$

for some $a \in \mathcal{O}_K$. As above, we may assume that α is real as $F/\mathbf{Q}$ is generalised dihedral. The ramification condition tells us that we may assume without loss of generality that $a = 2, 3, 6$, or 12. Now consider the $\mathcal{O}_K$-ideal $(31) = (2 + 9\sqrt{-3})(2 - 9\sqrt{-3})$, and notice that $\mathfrak{p} = (2 + 9\sqrt{-3})$ is trivial in the ring class group of conductor 6. We get

$$\sqrt[3]{a} = \mathrm{Frob}_{\mathfrak{p}}(\sqrt[3]{a}) \equiv \sqrt[3]{a}^{31} \equiv a^{10}\sqrt[3]{a} \pmod{\mathfrak{p}}.$$

This immediately rules out three of the four possibilities for a, and we conclude that $F = K(\sqrt[3]{2})$, whence we recover Euler's conjecture.

8 Exercises

1. Prove that -1 is a square in $\mathbf{F}_p$ for p an odd prime, if and only if $4 \mid p-1$, by considering the factorisation
$$x^{p-1}-1=(x^{(p-1)/2}-1)(x^{(p-1)/2}+1)$$
of polynomials over $\mathbf{F}_p$. Then give a similar proof of the following statements:
a) 2 is a square in $\mathbf{F}_p$ if and only if $16 \mid p^2-1$,
b) -3 is a square in $\mathbf{F}_p$ if and only if $3 \mid p-1$.

2. Give a Dedekind-style proof of the following results of Fermat:
$$\begin{cases} p=x^2+2y^2 & \Longleftrightarrow \quad p \equiv 1,3 \pmod 8 \\ p=x^2+3y^2 & \Longleftrightarrow \quad p \equiv 1 \pmod 3 \end{cases}$$
Can the methods of Euler or Heath–Brown be used to prove them?

3. Prove that a matrix $M \in \mathrm{PSL}_2(\mathbf{Z})$ is symmetric if and only if $M^{-1}=SMS$ where $S=\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

4. Show that a matrix $\gamma \in \mathrm{SL}_2(\mathbf{Z})$ acts on the coefficients of a quadratic form via a matrix $\Gamma \in \mathrm{SL}_3(\mathbf{Z})$, and the corresponding map
$$\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_3(\mathbf{Z}) : \gamma \mapsto \Gamma$$
is an injective group homomorphism.

5. Prove that an indefinite form $\langle a,b,c\rangle$ is reduced if and only if $ac<0$ and $b>|a+c|$.

6. Show that a binary quadratic form properly represents an integer n if and only if there is a form equivalent to it in which the coefficient of x^2 is n.

7. Prove that there are at most two reduced form of discriminant $\Delta<0$ representing a given prime p.

8. Show that a quadratic form represents 0 properly if and only if its discriminant is a square.

9. Recall that the number of *ambiguous* forms $\langle a,b,c\rangle \sim \langle a,-b,c\rangle$ is equal to the number of genera $\nu(\Delta)$. In this question, we will investigate *antiguous* forms $\langle a,b,c\rangle \sim \langle -a,b,-c\rangle$ and *reciprocal* forms $\langle a,b,c\rangle \sim \langle -a,-b,-c\rangle$.

   (a) Prove that $\langle -a,b,-c\rangle \sim \langle a,b,c\rangle$ for some form of discriminant Δ if and only if the same is true for every form of discriminant Δ if and only if $\langle -1,\sigma,\frac{\Delta-\sigma}{4}\rangle$ is in the principal class.

(b) Find and prove a characterisation of those discriminants Δ for which the form $\langle -1, \sigma, \frac{\Delta-\sigma}{4}\rangle$ is in the principal class, where $\sigma \in \{0, 1\}$ such that $4 \mid \Delta - \sigma$.

(c) Prove that $\langle -a, -b, -c\rangle \sim \langle a, b, c\rangle$ for some form of discriminant Δ if and only if the same is true for exactly $\nu(\Delta)$ forms of discriminant Δ if and only if $\langle -1, \sigma, \frac{\Delta-\sigma}{4}\rangle$ is in the principal genus.

(d) Find and prove a characterisation of those discriminants Δ for which the form $\langle -1, \sigma, \frac{\Delta-\sigma}{4}\rangle$ is in the principal genus, where $\sigma \in \{0, 1\}$ such that $4 \mid \Delta - \sigma$.

10. Find a discriminant Δ such that $|\mathrm{Cl}^+_\Delta| = 2|\mathrm{Cl}_\Delta|$ and $\mathrm{Cl}^+_\Delta \not\simeq \mathrm{Cl}_\Delta \times \mathrm{C}_2$. Characterise all such Δ.

11. Find all integral solutions to the following equations:

$$\begin{array}{lll}
a. & x^2 - 65y^2 & = & 7, \\
b. & x^2 - 82y^2 & = & 31, \\
c. & x^2 - 15y^2 & = & 34, \\
d. & x^2 - 67y^2 & = & 21. \\
e. & 3x^2 + xy - 5y^2 & = & 19. \\
f. & x^2 + xy + 7y^2 & = & 292.
\end{array}$$

12. Can you generalise Theorem 4.2 to forms which represent 0?

13. Find the class number of $\mathrm{Cl}^+(-163)$.

14. Prove that a reduced positive definite form $\langle a, b, c\rangle$ has order at most two in the class group if and only if $b = 0$, $a = b$, or $a = c$. Can you generalise this to indefinite forms?

15. (Euler) Let $f(x) = x^2 + x + 41$. Find the smallest non-negative integer n so that $f(n)$ is not prime. Use genus theory to explain why the answer is greater than $(163 - 7)/4$.

16. Suppose that $\langle 6, 2, 7\rangle$ represents m, and $\langle 2, 2, 21\rangle$ represents n. Show that $\langle 3, 2, 14\rangle$ represents mn.

17. Show that $\mathrm{Cl}^+(n^2) \simeq (\mathbf{Z}/n\mathbf{Z})^\times$.

18. Let $n \geq 3$ be an integer, and set $\Delta = -(2^n - 1)$. Show that $h^+(\Delta) \equiv 0 \pmod{n-2}$.
Hint: Consider the class of $\langle 2, 1, 2^{n-3}\rangle$.

19. Let $p = a^6 + 4b^6$ be a prime. Show that $\langle b^3, a^3, -b^3\rangle$ represents 1.

20. Use class field theory to prove an explicit criterion for when a prime is of the form $x^2 + 64y^2$.

Chapter II

Complex multiplication

We have now come to a theoretical resolution of the question when a prime is of the form $x^2 + ny^2$ for a general $n \geq 1$. We saw that we were able to prove Euler's conjecture for $n = 27$ by using Kummer theory to identify the relevant ring class field. However, in general this will not be possible. Take $n = 14$, for instance, then we are looking for an extension of degree 4 of the quadratic field $\mathbf{Q}(\sqrt{-14})$ which is unramified everywhere, but it is not at all clear how we could find this extension by hand. In general, the situation will be even worse.

We now turn to the question of whether we can make this criterion explicit for all n. In other words, can we generally compute explicit generators for all abelian extension of a number field? The Artin map is not sufficiently explicit for us to do this purely in terms of idèle class groups, so a new tool is needed. Drawing inspiration from the Kronecker–Weber theorem, which allows us to generate all abelian extensions of $\mathbf{Q}$ by adjoining roots of unity, which are precisely the torsion points of the algebraic group $\mathbf{G}_m$, we will see in this chapter that the class field theory of imaginary quadratic fields can likewise be made explicit by using instead the j-invariant and torsion points of different algebraic groups: Elliptic curves with complex multiplication. This will lead to a full answer of the question when a prime p is of the form $x^2 + ny^2$ when $n \geq 1$. We are left to wonder what happens when $n \leq 1$, or when we simply want to make the class field theory of a general number field explicit, and at present no such general theory is available.

9 Review of elliptic and modular functions

If $\tau \in \mathfrak{H}$ is a point in the upper half plane, we define the lattice $\Lambda_\tau = \mathbf{Z} + \mathbf{Z}\tau \subset \mathbf{C}$, and $E_\tau = \mathbf{C}/\Lambda_\tau$ the elliptic curve over $\mathbf{C}$ corresponding to τ. Many of the arithmetic properties of E_τ may be described using the Weierstrass functions. Most of the functions discussed in this section have been introduced in Prof. Darmon's course, so to avoid too much overlap, we will be brief and omit most proofs.

9.1 The Weierstrass functions.

We briefly recall the definition and basic properties of the functions $\sigma(z), \zeta(z)$, and $\wp(z)$.

The Weierstrass $\wp$-function. The elliptic curves $E)\tau$ have Weierstrass forms described via the *Weierstrass $\wp$-function*, which is defined by

$$\wp_T(z) = \frac{1}{z^2} + \sum_{\lambda \in \Lambda_\tau \backslash \{0\}} \left(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2} \right).$$

This defines a meromorphic function on the elliptic curve E_τ with a pole of order 2 at the identity, and no other poles. It has the following Laurent expansion around $z = 0$:

$$\wp_T(z) = \frac{1}{z^2} + \sum_{n \geq 1} (2n+1) G_{2n+2}(\tau) z^{2n},$$

where $G_{2n}(\tau)$ is the value at τ of the non-normalised Eisenstein series of weight $2n$ defined by

$$G_{2n}(\tau) = \sum_{\lambda \in \Lambda_\tau \backslash \{0\}} \frac{1}{\lambda^{2n}}.$$

The elliptic curve E_τ is isomorphic to the elliptic curve with Weierstrass equation

$$y^2 = 4x^3 - 60G_4(\Lambda_\tau)x - 140G_6(\Lambda_\tau),$$

via the map $z \in \mathbf{C}/\Lambda_\tau \mapsto (\wp_\tau(z), \wp'_\tau(z), 1)$. Conversely, every elliptic curve over $\mathbf{C}$ arises this way. This is called the uniformisation theorem for elliptic curves, and underscores the importance of the Weierstrass $\wp$-function: It connects the complex analytic theory of elliptic curves to the algebraic world, by describing explicitly a Weierstrass model for the corresponding elliptic curve.

The Weierstrass ζ-function. As can be seen from the Laurent expansion of the Weierstrass $\wp$-function, it has double poles with vanishing residues at the points of Λ_τ. This means we can integrate this function to obtain a new one:

Theorem 9.1. *The power series*

$$\zeta_T(z) = \frac{1}{z} + \sum_{\lambda \in \Lambda_\tau \backslash \{0\}} \left(\frac{1}{z-\lambda} + \frac{1}{\lambda} + \frac{z}{\lambda^2} \right)$$

is absolutely and uniformly convergent on compact open subsets of $\mathbf{C} \backslash \Lambda_\tau$*, and satisfies*

- $\frac{d}{dz}\zeta_\tau(z) = -\wp_T(z)$*,*
- $\zeta_\tau(z+\lambda) = \zeta_\tau(z) + \eta_\tau(\lambda)$*, where* $\eta : \Lambda_\tau \to \mathbf{C}$ *is a homomorphism,*

- $\tau\eta_\tau(1) - \eta_\tau(\tau) = 2\pi i.$

The additive homomorphism $\eta_\tau : \Lambda_\tau \to \mathbf{C}$ is called the *quasi-period* of $\zeta_\tau(z)$. It measures the failure of ζ_τ to be an elliptic function, and encodes the periods of differentials on E_τ when paired with a basis of singular homology, see below. From the Laurent series expansion of $\wp$, we deduce that the Laurent series expansion around $z = 0$ of the ζ-function is

$$\zeta_\tau(z) = \frac{1}{z} - \sum_{n\geq 1} G_{2n+2}(\tau) z^{2n+1}.$$

The Weierstrass σ-function. We now come to the Weierstrass σ-function, which we obtain morally from integrating and exponentiating the Weierstrass ζ-function. More precisely:

Theorem 9.2. *The infinite product*

$$\sigma_\tau(z) = z \prod_{\lambda\in\Lambda_\tau\setminus\{0\}} \left(1 - \frac{z}{\lambda}\right) \exp\left(\frac{z}{\lambda} + \frac{z^2}{2\lambda^2}\right)$$

converges absolutely to a holomorphic function on $\mathbf{C}$ *with simple zeroes on* Λ*, and no zeroes elsewhere.*

The Weierstrass σ-function satisfies the following transformation formula for the lattice Λ_τ :

$$\sigma_\tau(z+\lambda) = \sigma_\tau(z)\cdot\psi(\lambda)e^{\eta_\tau(\lambda)(z+\frac{\lambda}{2})}.$$

Here, $\eta_\tau : \Lambda_\tau \to \mathbf{C}$ is the *quasi-period* of $\zeta_\tau(z)$, and $\psi(\lambda)$ is 1 if $\lambda \in 2\Lambda_\tau$ and -1 otherwise. This shows that σ is not periodic, but it is up to some constant. We see that the Weierstrass functions are all related by

$$\text{dlog}\,\sigma_\tau(z) = \zeta_\tau(z), \qquad -\frac{d\zeta_\tau}{dz}(z) = \wp_\tau(z).$$

There is a beautiful product formula relating the Weierstrass $\wp$ and σ functions.

Theorem 9.3. *For* $z_1, z_2 \notin \Lambda_\tau$*, we have the identity*

$$\wp_\tau(z_1) - \wp_\tau(z_2) = -\frac{\sigma_\tau(z_1+z_2)\sigma_\tau(z_1-z_2)}{\sigma_\tau(z_1)^2\sigma_\tau(z_2)^2}. \tag{II.1}$$

Proof. Fix z_2, and consider both sides of the identity above as functions of z_1. Then both sides define even elliptic functions with double poles at the points of Λ_τ, and zeroes at $\pm z_2 + \Lambda_\tau$. This means that both functions must differ by a multiplicative constant, which we may find by multiplying both sides by z and comparing residues at $z = 0$. □

9.2 The j-function.

A central role will be played by the j-function, which is an $\mathrm{SL}_2(\mathbf{Z})$-invariant holomorphic function on $\mathfrak{H}$. Before we define it, let us recall for the sake of completeness the definition of the Dedekind η-function

$$\eta(\tau) = q^{1/24}\prod_{n\geq 1}(1-q^n),$$

where $\tau \in \mathfrak{H}$ and as usual $q = e^{2\pi i\tau}$. This is a holomorphic function on $\mathfrak{H}$ which satisfies the following transformation laws:

$$\eta(\tau+1) = \zeta_{24}\eta(\tau), \qquad \eta(-1/\tau) = \sqrt{-i\tau}\eta(\tau),$$

where we chose the branch of the square root which is positive on the positive real axis. Of course, these two transformation laws imply a transformation law for the entire group $\mathrm{SL}_2(\mathbf{Z})$ by induction on the length of the word. The expression one ends up getting involves so called *Dedekind sums*, which are elementary, albeit slightly mystifying, expressions in terms of the matrix entries.

From the Dedekind η-function, we obtain the *modular discriminant* $\Delta(\tau) = \eta(\tau)^{24}$, which by the above is clearly a modular form of weight 12 on $\mathrm{SL}_2(\mathbf{Z})$. This is a remarkable function, which gets its name from the fact that $\Delta(\tau)$ is precisely the discriminant of the elliptic curve E_τ in the Weierstrass form specified above. It is non-vanishing on $\mathfrak{H}$, and has a simple zero at the cusps. We define the *j-invariant* to be the function

$$j(\tau) = \frac{E_4^3(\tau)}{\Delta(\tau)} = \frac{1}{q} + 744 + 196884q + 21493760q^2 + \dots,$$

where

$$E_4(q) = 1 + 240\sum_{n\geq 1}\sigma_3(n)q^n, \qquad \sigma_3(n) = \sum_{d|n} d^3$$

is the normalised weight 4 Eisenstein series. Clearly, $j(z)$ is a holomorphic function on $\mathfrak{H}$ which has meromorphic continuation to the cusps, and is $\mathrm{SL}_2(\mathbf{Z})$-invariant. The j-invariant of E_τ is equal to $j(\tau)$, the value of the j-function at τ. The values of the j-function therefore determine the isomorphism type of E_τ over $\mathbf{C}$.

The j-function is not an algebraic function in the sense that it is not true that $j(\tau)^\sigma = j(\tau^\sigma)$ when $\sigma \in \mathrm{Aut}(\mathbf{C})$. However, as we will see in this chapter, such a statement is true when τ is a quadratic irrationality. The value $j(\tau)$ in that case is algebraic, and we can even be very specific about the extension of $\mathbf{Q}$ it generates. This is more or less the content of CM theory!

9.3 The Weber functions.

We will need one more class of modular functions, which was not discussed in Prof. Darmon's class. First, we need a cube root of the j-invariant. The modular discriminant does not vanish on the upper half plane, so we may choose a holomorphic cube root which is real-valued on the imaginary axis. This allows us to define

$$\gamma_2(q) = \frac{E_4(q)}{\Delta(q)^{1/3}},$$

which is a holomorphic function on $\mathfrak{H}$. It transforms as follows:

$$\begin{cases} \gamma_2(z+1) &= \zeta_3^2\gamma_2(z), \\ \gamma_2(\frac{-1}{z}) &= \gamma_2(z). \end{cases}$$

Both of these identities are easy to prove: They clearly hold up to roots of unity. As γ_2 is real along the imaginary axis, the second identity follows immediately. For the first, write

$$\gamma_2(q) = q^{-1/3}\left(1 + 744q + 196884q^2 + \dots\right)^{1/3},$$

where the binomial expansion of the second factor is clearly a power series over $\mathbf{Q}$, defining a holomorphic function invariant under $z \mapsto z+1$, which implies the above identity. By an easy induction on the word length of a general element in $\mathrm{SL}_2(\mathbf{Z})$, we can show that

$$\gamma_2\left(\frac{az+b}{cz+d}\right) = \zeta_3^{ac-ab+a^2cd-cd}\gamma_2(z),$$

from which we see that γ_2 is a modular function for the congruence subgroup

$$\Gamma_0^0(3)^+ = \left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \ : \ a \equiv d \equiv 0 \pmod 3 \text{ or } b \equiv c \equiv 0 \pmod 3\right\}.$$

The *Weber functions* $\mathfrak{f}, \mathfrak{f}_1, \mathfrak{f}_2$ are defined to be

$$\begin{aligned} \mathfrak{f}(z) &= \zeta_{48}^{-1}\frac{\eta\left(\frac{z+1}{2}\right)}{\eta(z)} &&= q^{-1/48}\prod_{n\geq 1}(1+q^{n-\frac{1}{2}}), \\ \mathfrak{f}_1(z) &= \frac{\eta\left(\frac{z}{2}\right)}{\eta(z)} &&= q^{-1/48}\prod_{n\geq 1}(1-q^{n-\frac{1}{2}}), \\ \mathfrak{f}_2(z) &= \sqrt{2}\frac{\eta(2z)}{\eta(z)} &&= \sqrt{2}q^{1/24}\prod_{n\geq 1}(1+q^{n}). \end{aligned}$$

These functions satisfy various identities. For instance, it follows straight from the definition that

$$\mathfrak{f}_1(2z)\mathfrak{f}_2(z) = \mathfrak{f}(z)\mathfrak{f}_1(z)\mathfrak{f}_2(z) = \sqrt{2}.$$

The definition of the Weber functions may seem quite unnatural at first, but these functions come up organically when we talk about the 2-torsion on the universal elliptic curve over $\mathbf{C}$. More precisely

Theorem 9.4. *Let* $e_1 = \wp_\tau(\tau/2), e_2 = \wp_\tau(1/2)$, *and* $e_3 = \wp_\tau((\tau+1)/2)$, *then we have the identities*

$$\begin{cases} e_2 - e_1 &= \pi^2\eta(\tau)^4\mathfrak{f}(\tau)^8, \\ e_2 - e_3 &= \pi^2\eta(\tau)^4\mathfrak{f}_1(\tau)^8, \\ e_3 - e_2 &= \pi^2\eta(\tau)^4\mathfrak{f}_2(\tau)^8. \end{cases}$$

Proof. By the product formula (II.1) we have that

$$e_2 - e_1 = e^{\eta_\tau(\tau)/2}\frac{\sigma_\tau((\tau+1)/2))^2}{\sigma_\tau(\tau/2)^2\sigma_\tau(1/2)^2},$$

and similarly for the other identities. The Weber functions have a convenient product formula for their q-expansions, so the result would be a formal manipulation if we could find a good product formula for the q-expansion of σ. In fact, we can show that

$$\sigma_\tau(z) = (2\pi i)^{-1}e^{\eta_\tau(1)z^2/2}(q^{1/2} - q^{-1/2})\prod_{n\geq 1}\frac{(1-q_\tau^n q)(1-q_\tau^n/q)}{(1-q_\tau^n)^2},$$

where $\eta_\tau(1)$ is the value of the quasi-period of ζ_τ at 1, and $q_\tau = e^{2\pi i\tau}$. To prove this, it suffices to take the quotient of both sides, and note that this yields by the transformation laws of $\sigma_\tau(z)$ an elliptic function for Λ_τ. This elliptic function has no zeros, and is hence constant. By letting $z \to 0$, we recover that the constant is 1. □

These identities have a pleasant corollary that will be of practical use for us later when we compute singular moduli.

Corollary 9.1. *The following identity holds:*

$$\gamma_2 = \frac{\mathfrak{f}^{24}-16}{\mathfrak{f}^8} = \frac{\mathfrak{f}_1^{24}+16}{\mathfrak{f}_1^8} = \frac{\mathfrak{f}_2^{24}+16}{\mathfrak{f}_2^8}.$$

Proof. Choose $\tau \in \mathfrak{H}$, and consider the numbers e_1, e_2, e_3 as above, which are the roots of the polynomial

$$4x^3 - 60G_4(\tau)x - 140G_6(\tau).$$

This implies that $e_1 + e_2 + e_3 = 0$, and also that $15G_4(\tau) = -(e_1e_2 + e_2e_3 + e_1e_3)$. Note that the roots summing up to 0 implies that $-3(e_1e_2 + e_2e_3 + e_1e_3) = (e_2 - e_1)^2 - (e_2 - e_3)(e_3 - e_1)$, from which we obtain

$$45G_4(\tau) = \pi^4E_4(\tau) = -3((e_2-e_1)^2 - (e_2-e_3)(e_3-e_1).,$$

Now use Theorem 9.4 to deduce that $\pi^4E_4(\tau) = \pi^4\eta^8(\tau)\mathfrak{f}^{16}(\tau) - \pi^4\eta^8(\tau)\mathfrak{f}_1^8(\tau)\mathfrak{f}_2^8(\tau)$, so that the statement follows from the identity $\mathfrak{f}(\tau)\mathfrak{f}_1(\tau)\mathfrak{f}_2(\tau) = \sqrt{2}$. □

Finally, it is a formal consequence of the transformation laws of the Dedekind η-function, and the definition of the Weber functions, that we have the identities:

$$\left\|\begin{array}{lcl} \mathfrak{f}(z+1) & = & \zeta_{48}^{-1}\mathfrak{f}_1(z) \\ \mathfrak{f}_1(z+1) & = & \zeta_{48}^{-1}\mathfrak{f}(z) \\ \mathfrak{f}_2(z+1) & = & \zeta_{24}\mathfrak{f}_2(z) \end{array}\right. \quad \left\|\begin{array}{lcl} \mathfrak{f}(-1/z) & = & \mathfrak{f}(z) \\ \mathfrak{f}_1(-1/z) & = & \mathfrak{f}_2(z) \\ \mathfrak{f}_2(-1/z) & = & \mathfrak{f}_1(z) \end{array}\right.$$

It follows that $\mathfrak{f}^6$ is a modular function for the congruence subgroup

$$\Gamma_0^0(8) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \ : \ b \equiv c \equiv 0 \pmod 8 \right\},$$

whereas $\mathfrak{f}_2$ is a modular function for $\Gamma(24)$.

9.4 The geometric interpretation.

All of the above elliptic functions have a clear description in geometric terms on the corresponding elliptic curve. As we saw above, the isomorphism

$$\mathbf{C}/\Lambda_\tau \to E_\tau : z \mapsto (\wp(z) : \wp'(z) : 1)$$

describes how the algebraic description of E_τ, as a curve with a Weierstrass equation, connects to the complex analytic description $E_\tau = \mathbf{C}/\Lambda_\tau$. This shows that the standard basis

$$\mathrm{H}^1_{\mathrm{dR}}(E_\tau/\mathbf{C}) = \langle \frac{dx}{y}, x\frac{dx}{y} \rangle$$

admits a simple description on the analytic side, as dz and $\wp(z)dz$. Given an element $\lambda \in \Lambda_\tau$, we obtain a corresponding element in the singular homology $\mathrm{H}_1(E/\mathbf{C})$ by considering the image of a path from 0 to λ in $\mathbf{C}$. The Poincaré duality pairing may now be computed explicitly by computing the line integrals

$$\int_0^1 dz = 1, \int_0^\tau dz = \tau, \text{ and } \int_0^1 \wp(z)dz, \int_0^\tau \wp(z)dz.$$

The Weierstrass ζ-function provides us with a primitive for the form $\wp(z)dz$, as $-\zeta'(z) = \wp(z)$, so we see that the periods are described precisely by the failure of $\zeta(z)$ to be elliptic, and are encoded in the quasi-period map defined above.

The Weierstrass σ-function also has arithmetic significance, though it lies much deeper. The differential $\zeta(z)dz$ is genuinely invariant under Λ_τ, and hence descends to E_τ. It is a differential with a simple pole at the origin, and the residue is an integer. Such differentials are called *differentials of the third kind*, and come up in the theory of height functions. This is the conceptual habitat of $\sigma(z)$, and its values acquire significance when viewed in that context. We will not discuss this further in this course, but it is a fascinating and important part of arithmetic geometry.

10 Complex multiplication for elliptic curves

We say that an elliptic curve $E_{\mathbf{C}}$ has *complex multiplication*, or we simply say E is a CM elliptic curve, if $\mathrm{End}(E) \neq \mathbf{Z}$. Thanks to complex uniformisation of elliptic curves, this translates into a very nice condition as follows: Suppose E_{τ_1} and E_{τ_2} are two elliptic curves over $\mathbf{C}$, attached to the points $\tau_1, \tau_2 \in \mathfrak{H}$. Every complex analytic homomorphism arises from a linear map $\mathbf{C} \to \mathbf{C} : z \mapsto \mu z$ such that $\mu\Lambda_{\tau_1} \subseteq \Lambda_{\tau_2}$. More precisely, we have

$$\mathrm{Hom}(E_{\tau_1}, E_{\tau_2}) \simeq \{\mu \in \mathbf{C} : \mu\Lambda_{\tau_1} \subseteq \lambda_{\tau_2}\}.$$

As a consequence, we obtain the following lemma:

Lemma 10.1. *The elliptic curves E_{τ_1} and E_{τ_2} are isogenous, resp. isomorphic, if and only if there exists an $\alpha \in \mathrm{GL}_2^+\mathbf{Q}$, resp. $\mathrm{SL}_2\mathbf{Z}$, such that $\alpha \cdot \tau_1 = \tau_2$.*

Proof. Suppose first that E_{τ_1} and E_{τ_2} are isogenous, then there is a $\mu \in \mathbf{C}$ so that

$$\mu \begin{pmatrix} \tau_1 \\ 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \tau_2 \\ 1 \end{pmatrix} \qquad \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2\mathbf{Q},$$

where $a, b, c, d \in \mathbf{Z}$. Since $\tau_1, \tau_2 \in \mathfrak{H}$, we must have $\det \alpha > 0$, as required. Note that μ defines an isomorphism if and only if $\alpha \in \mathrm{SL}_2(\mathbf{Z})$. Conversely, suppose that such a matrix α exists, then set $\mu = c\tau_2 + d$. We easily check that μ defines the required isogeny, which is an isomorphism if and only if $\alpha \in \mathrm{SL}_2(\mathbf{Z})$. □

Over $\mathbf{C}$, a result of Deuring (see Appendix B) assures us that $\mathrm{End}_{\mathbf{Q}}(E)$ is either $\mathbf{Q}$, or an imaginary quadratic field. This may easily be deduced from the lemma we just proved, as the existence of a non-scalar matrix $\alpha \in \mathrm{GL}_2^+\mathbf{Q}$ fixing $\tau \in \mathfrak{H}$ is equivalent to τ satisfying an irreducible quadratic polynomial over $\mathbf{Q}$. We conclude that E_τ is a CM elliptic curve if and only if $\mathbf{Q}(\tau)$ is an imaginary quadratic field K, in which case we have

$$\mathrm{End}_{\mathbf{Q}}(E) \simeq \mathbf{Q}(\tau) = K.$$

We say in this case that E has CM by K, or if we want to be even more precise, by the order $\mathrm{End}(E)$ in K. There are two possible isomorphisms $\mathrm{End}_{\mathbf{Q}}(E) \simeq K$, and it will frequently be useful to normalise the isomorphism appropriately. We say the isomorphism is *normalised* if the endomorphism corresponding to $\mu \in K$ has the property that it multiplies the invariant differential by μ, as opposed to $\bar{\mu}$. In other words, if

$$[\mu]^*\omega = \mu\omega, \qquad \text{where } \omega = dz.$$

This normalisation is convenient for many reasons, for instance if E_1 and E_2 are elliptic curves with CM by the same quadratic order, and we normalise the isomorphisms, then every isogeny $\varphi \in \mathrm{Hom}(E_1, E_2)$ has the property that

$$\varphi \circ [\mu]_{E_1} = [\mu]_{E_2} \circ \varphi.$$

This is easy to show on the cotangent spaces at the origin, which implies the above equality on the elliptic curves in the usual way. Now let us fix a quadratic imaginary field K, and classify all elliptic curves with CM by K.

Theorem 10.1. *Let E be an elliptic curve with CM by an order $\mathcal{O} = \operatorname{End}(E)$ in K, then $E \simeq \mathbf{C}/\mathfrak{a}$ for $\mathfrak{a}$ a proper ideal in $\mathcal{O}$, and conversely every curve of that form has endomorphism ring isomorphic to $\mathcal{O}$. This induces a bijection*

$$\{\textit{Isomorphism classes of } E \textit{ s.t. } \operatorname{End}(E) \simeq \mathcal{O}\} \longrightarrow \operatorname{Cl}(\mathcal{O}).$$

Proof. If E_τ is an elliptic curve with $\operatorname{End}(E) \simeq \mathcal{O}$, then by the above the discussion we must have $\mathcal{O} = \{\mu \in \mathbf{C} : \mu\Lambda_\tau \subseteq \Lambda_\tau\}$, which means that $\Lambda_\tau = \mathfrak{a}$ is a proper ideal of $\mathcal{O}$. This defines a map from the set of all such elliptic curves, to the class group $\operatorname{Cl}(\mathcal{O})$. If $E = \mathbf{C}/\mathfrak{a}$ and $E' = \mathbf{C}/\mathfrak{b}$ are two such curves that are isomorphic, then there exists a $\mu \in \mathbf{C}^\times$ such that $\mu\mathfrak{a} = \mathfrak{b}$. This implies that in fact $\mu \in K^\times$, and hence that $\mathfrak{a}$ and $\mathfrak{b}$ are in the same class. Conversely, if $\mathfrak{a}$ is a proper ideal of $\mathcal{O}$, then clearly $E = \mathbf{C}/\mathfrak{a}$ is an elliptic curve with endomorphism ring $\mathcal{O}$, which finishes the proof. □

So far, we have only used complex analytic arguments. Already, the results we have obtained imply some arithmetic statements. To get statements over subfields of $\mathbf{C}$, we bring Galois actions into play. If E is an elliptic curve over $\mathbf{C}$, it has a Weierstrass equation provided by the uniformisation theorem for elliptic curves. For any $\sigma \in \operatorname{Aut}(\mathbf{C})$, we denote E^σ for the elliptic curve whose Weierstrass equation is obtained by letting σ act on the coefficients of the Weierstrass equation of E. We will deduce a number of algebraicity statements from the following heuristic:

> Let k be a subfield of $\mathbf{C}$. If $X^\sigma = X$ for all $\sigma \in \operatorname{Aut}(\mathbf{C}/k)$, then X is rational over k.

The statement above is merely a heuristic, and it applies in a variety of different situations, which is why the nature of X is deliberately kept vague. To see this principle in action, let us show that the fields of definition of various objects related to elliptic curves actually descend to an arithmetic base field.

Lemma 10.2. *Let E_1, E_2 be elliptic curves defined over a subfield k of $\mathbf{C}$, and let $\overline{k}$ be the algebraic closure of k in $\mathbf{C}$. Then every element of $\operatorname{Hom}(E_1, E_2)$ is defined over $\overline{k}$. Moreover, if $\operatorname{End}(E_1) = \mathbf{Z}$, then $\lambda^\sigma = \pm\lambda$ for any $\sigma \in \operatorname{Aut}(\overline{k}/k)$ and $\lambda \in \operatorname{Hom}(E_1, E_2)$.*

Proof. If $\lambda \in \operatorname{Hom}(E_1, E_2)$ and $\sigma \in \operatorname{Aut}(\mathbf{C}/k)$, then also $\lambda^\sigma \in \operatorname{Hom}(E_1, E_2)$. The set $\operatorname{Hom}(E_1, E_2)$ is countable, as it is identified with the set of complex numbers that send the lattice Λ_{τ_1} into Λ_{τ_2}, where $\tau_1, \tau_2 \in \mathfrak{H}$ correspond to E_1, E_2. This means that λ must be defined over $\overline{k}$. Finally, if $\operatorname{End}(E_1) = \mathbf{Z}$ then also $\operatorname{Hom}(E_1, E_2) \simeq \mathbf{Z}$, and it follows that $m\lambda^\sigma = n\lambda$ for some $m, n \in \mathbf{Z}$. Since the degrees of λ and λ^σ are equal, we see that $m = \pm n$ as required. □

As another illustration, we show that endomorphisms of CM elliptic curves are always defined over a quadratic extension of the base field. More precisely:

Lemma 10.3. *Let E be an elliptic curve defined over a subfield k of $\mathbf{C}$, with CM by K. Then every element of $\operatorname{End}(E)$ is defined over kK.*

Proof. Let ω be the invariant differential corresponding to dz on $\mathbf{C}$, and normalise the isomorphism $\operatorname{End}_{\mathbf{Q}}(E) \simeq K$. As ω and is defined over k, we compute

$$([\mu]^\sigma)^*\omega = ([\mu]^*\omega)^\sigma = (\mu\omega)^\sigma = \mu\omega = [\mu]^*\omega,$$

for all $\sigma \in \operatorname{Aut}(\mathbf{C}/Kk)$. This shows that $[\mu]$ is defined over Kk. □

Finally, we come to an important result, which will be superseded by much stronger results in the following sections. However, it is interesting to note that the mere algebraicity of singular moduli is very easy to prove from elementary principles.

Lemma 10.4. *Let E_τ be an elliptic curve with CM by K. Then $j(\tau)$ is algebraic.*

Proof. For any $\sigma \in \operatorname{Aut}(\mathbf{C})$, the ring $\operatorname{End}_{\mathbf{Q}}(E^\sigma)$ is isomorphic to K. It must therefore correspond, by Theorem 10.1, to a proper ideal class in some order $\mathcal{O}$ of K. All such orders are indexed by their conductor, which is a natural number, and for a fixed order there are finitely many proper ideal classes. As we have $j(E^\sigma) = j(E)^\sigma$, the set

$$\{j(E)^\sigma \;:\; \sigma \in \operatorname{Aut}\mathbf{C}\}$$

is countable, so that $j(E)$ must be algebraic. □

The above lemma was quite easy to prove. But we can be much more specific about the nature of the algebraic number $j(E)$. We will see later that it is in fact an algebraic integer, and that we can be very specific about the extension it generates over K. The latter is the content of the main theorem of complex multiplication, and it lies much deeper.

11 The main theorem of complex multiplication

We now come to the main theorem of complex multiplication, which fine-tunes our understanding of the set $j(E)^\sigma$, for E an elliptic curve with CM by K, and $\sigma \in \operatorname{Aut}(\mathbf{C}/K)$.

The Artin map. We will use the adelic language of class field theory as summarised in the appendix. Most important in what follows is the Artin map $\varphi : C_K \to \operatorname{Gal}(K^{\mathrm{ab}}/K)$ which is surjective. More precisely, there is an exact sequence

$$1 \to \overline{K^\times \mathbf{A}_{K,+}^{f,\times}} \to \mathbf{A}_K^\times \xrightarrow{\varphi_K} \operatorname{Gal}(K^{\mathrm{ab}}/K) \to 1,$$

where $\mathbf{A}_{K,+}^{f,\times}$ is the connected component of the identity in the archimedean part $\mathbf{A}_K^{f,\times}$ of the idèles, and the overline denotes the topological closure. In fact, when K is an imaginary quadratic field, as will be the case in our applications, then $K^\times \mathbf{A}_{K,+}^{f,\times}$ is already closed. Indeed, $K^\times = \mathcal{O}_K^\times$ is finite, and the archimedean part is isomorphic to $\mathbf{C}^\times$.

Let us first fix some notation for this section. Let E be an elliptic curve with CM by K, where we have normalised the isomorphism $\mathrm{End}_{\mathbf{Q}}(E) \simeq K$. We know that there exists an isomorphism

$$\xi : \mathbf{C}/\mathfrak{a} \xrightarrow{\sim} E$$

for some lattice $\mathfrak{a}$ in K. Let p be a rational prime, and define $K_p = K \otimes_{\mathbf{Q}} \mathbf{Q}_p$. Then $\mathfrak{a}_p = \mathfrak{a} \otimes_{\mathbf{Z}} \mathbf{Z}_p$ is a lattice in K_p. We now recall how to multiply a lattice by an idèle of K:

Lemma 11.1. *Let $x \in \mathbf{A}_K^\times$, then there exists a lattice $x\mathfrak{a}$ in K such that $(x\mathfrak{a})_p = x_p\mathfrak{a}_p$ for all p, and we have a commutative diagram, where the horizontal maps are isomorphisms*

$$\begin{tikzcd}
K_p/\mathfrak{a}_p \arrow[r, "x_p"] \arrow[d] & K_p/x_p\mathfrak{a}_p \arrow[d] \\
K/\mathfrak{a} \arrow[r, "x"] & K/x\mathfrak{a}
\end{tikzcd}$$

Proof. Recall that to the idèle x, we can naturally attach the fractional ideal $\mathfrak{x}$ of K defined by

$$\mathfrak{x} = \prod_{\mathfrak{p} \text{ finite}} \mathfrak{p}^{v_{\mathfrak{p}}(x)}.$$

In general, whenever L is a lattice in K and $\mathfrak{i}$ is a fractional ideal of K, we can define

$$\mathfrak{i}L = \{i_1 l_1 + \ldots + i_n l_n \ : \ i_i \in \mathfrak{i}, \ l_i \in L\}$$

which is another lattice in K. Applying this to $\mathfrak{i} = \mathfrak{x}$ and $L = \mathfrak{a}$, we obtain the required lattice. To define the horizontal multiplication by x map, note that $K/\mathfrak{a}$ is a torsion $\mathbf{Z}$-module, and as such it is the direct sum of its p-primary components. In other words, there is a canonical isomorphism

$$K/\mathfrak{a} \simeq \bigoplus_p K_p/\mathfrak{a}_p,$$

so that we may simply define the multiplication by x map locally as multiplication by x_p. □

Remark. The above decomposition of $K/\mathfrak{a}$ into local pieces is taken from Shimura [Shi70, Section 5.2], and there is a much more general version for torsion modules over a Dedekind domain. A very precise description is given in Silverman [Sil09, Section II.8].

This gives us a method to modify a lattice $\mathfrak{a}$ in K by an idèle. The main theorem of complex multiplication says that this is precisely the modification needed to obtain the lattice corresponding to the image of E under the Galois action associated with the idèle via class field theory.

Theorem 11.1. *Let $\sigma \in \mathrm{Aut}(\mathbf{C}/K)$, and $s \in \mathbf{A}_K^\times$ an element corresponding to $\sigma \mid_{K^{ab}}$ via the Artin map attached to K. There exists an isomorphism $\xi_s : \mathbf{C}/s^{-1}\mathfrak{a} \longrightarrow E^\sigma$ such that the following diagram commutes:*

$$\begin{tikzcd}
K/\mathfrak{a} \arrow[r, "\xi"] \arrow[d, "s^{-1}"'] & E \arrow[d, "\sigma"] \\
K/s^{-1}\mathfrak{a} \arrow[r, "\xi_s"] & E^\sigma
\end{tikzcd}$$

Proof. This proof is quite long, so we will present a sketch of the arguments in Shimura [Shi70, Theorem 5.4] by breaking it up into several steps.

Step 1. First, we note that if we prove the statement for a given elliptic curve E, it will also hold for any elliptic curve isomorphic to it. Indeed, if $i : E' \overset{\sim}{\to} E$ is an isomorphism, and the statement of the theorem holds for E', then (with obvious notation) we have that $\mathfrak{a}' = \gamma\mathfrak{a}$ for some $\gamma \in K^\times$, and we have the following commutative diagram:

$$\begin{tikzcd}
K/\mathfrak{a} \arrow[r, "\gamma"] \arrow[d, "s^{-1}"'] & K/\mathfrak{a}' \arrow[r, "\xi'"] \arrow[d, "s^{-1}"'] & E' \arrow[r, "i"] \arrow[d, "\sigma"'] & E \arrow[d, "\sigma"'] \\
K/s^{-1}\mathfrak{a} \arrow[r, "\gamma"] & K/s^{-1}\mathfrak{a}' \arrow[r, "\xi'_s"] & (E')^\sigma \arrow[r, "i^\sigma"] & E^\sigma
\end{tikzcd}$$

Now if we set $\xi_s(z) = i^\sigma \circ \xi'_s(\gamma z)$ then we see that the statement of the theorem for E follows from that for E'. Henceforth, we may therefore assume that E is actually defined over $K(j(E))$, and likewise for any other elliptic curve we consider.

Step 2. Now we reduce the statement to the case where $\mathfrak{a}$ is a fractional ideal, or in other words the case $\mathrm{End}(E) \simeq \mathcal{O}_K$. Choose a fractional ideal $\mathfrak{b}$ in K that is contained in $\mathfrak{a}$, and let E' be an elliptic curve with isomorphism $\xi' : \mathbf{C}/\mathfrak{b} \to E'$, and let $\lambda : E' \to E$ be the isogeny obtained from the inclusion $\mathfrak{b} \subseteq \mathfrak{a}$. Assume that the statement of the theorem holds for E'. We have that $\mathrm{Ker}(\lambda) = \xi'(\mathfrak{a}/\mathfrak{b})$, and hence

$$\mathrm{Ker}(\lambda^\sigma) = \mathrm{Ker}(\lambda)^\sigma = \xi'(\mathfrak{a}/\mathfrak{b})^\sigma = \xi'_s(s^{-1}\mathfrak{a}/s^{-1}\mathfrak{b}).$$

Now let $\lambda' : (E')^\sigma \to E''$ be the isogeny obtained from the inclusion $s^{-1}\mathfrak{b} \subseteq s^{-1}\mathfrak{a}$, where we fix a corresponding isomorphism $\eta : \mathbf{C}/s^{-1}\mathfrak{a} \to E''$. We easily see that $\mathrm{Ker}(\lambda') = \xi'_s(s^{-1}\mathfrak{a}/s^{-1}\mathfrak{b}) = \mathrm{Ker}(\lambda^\sigma)$. This means that we may find an isomorphism $\psi : E'' \to E^\sigma$ so that $\psi \circ \lambda' = \lambda^\sigma$. This yields a commutative diagram

$$\begin{tikzcd}
\mathbf{C}/s^{-1}\mathfrak{b} \arrow[r, "\xi'_s"] \arrow[d, "s^{-1}"'] & (E')^\sigma \arrow[d, "\sigma"] \\
\mathbf{C}/s^{-1}\mathfrak{a} \arrow[r, "\psi \circ \eta"] & E^\sigma
\end{tikzcd}$$

Setting $\xi_s = \psi \circ \eta$, we calculate that for any $u \in K$ we have

$$\xi(u)^\sigma = \lambda^\sigma(\xi'(u)^\sigma) = \lambda^\sigma(\xi'_s(s^{-1}u)) = \xi_s(s^{-1}u).$$

This shows that the statement of the theorem holds for E if it holds for E', so we may assume henceforth that $\mathcal{O}$ is the maximal order in K.

Step 3. Starting from σ, we will now construct a suitable lattice. We assume that $\mathrm{End}(E) \simeq \mathcal{O}_K$, and number the j-invariants of a set of representatives $E_1, E_2, \ldots, E_h$ for the elliptic curves with CM by $\mathcal{O}_K$ as $j_1, j_2, \ldots, j_h$. Let m be a large enough natural number (see later). Now choose a finite Galois extension L/K such that the following conditions are satisfied:

- L contains the ray class field $K_{(m)}$,
- L contains the coordinates of all points in $E[m]$,
- L contains $j_1, j_2, \ldots, j_h$.

Once such an L is fixed, choose a prime ideal $\mathfrak{q}$ in L such that the following conditions are satisfied:

- $\mathfrak{q}$ lies above a prime p in $\mathbf{Q}$, which is split in K and unramified in L,
- $\sigma\, |_L = \mathrm{Frob}_\mathfrak{q} \in \mathrm{Gal}(L/K)$,
- $\mathfrak{q} \nmid (6m)$,
- The curves E_i^τ have good reduction modulo $\mathfrak{q}$, $^\forall \tau \in \mathrm{Gal}(L/K)$,
- The residue classes of $j_1, j_2, \ldots, j_h$ are different modulo $\mathfrak{q}$.

Note that the last three conditions are guaranteed by choosing $\mathfrak{q}$ of large enough norm, and by Chebotarev density we can ensure the first two conditions. Once these choices have been made, we obtain an elliptic curve $\mathbf{C}/\mathfrak{p}^{-1}\mathfrak{a}$. This curve still has CM by $\mathcal{O}_K$, so that we may find an isomorphism $\eta : \mathbf{C}/\mathfrak{p}^{-1}\mathfrak{a} \to E_i$ for some i. Choosing an integral ideal $\mathfrak{x}$ in K prime to p such that $\mathfrak{x}\mathfrak{p} = \alpha\mathcal{O}_K$, we obtain a commutative diagram

$$\begin{tikzcd}
\mathbf{C} \arrow[r] \arrow[d, "\mathrm{Id}"'] & \mathbf{C}/\mathfrak{a} \arrow[r, "\xi"] \arrow[d] & E \arrow[d, "\lambda"] \\
\mathbf{C} \arrow[r] \arrow[d, "\alpha"'] & \mathbf{C}/\mathfrak{p}^{-1}\mathfrak{a} \arrow[r, "\eta"] \arrow[d] & E_i \arrow[d, "\mu"] \\
\mathbf{C} \arrow[r] & \mathbf{C}/\mathfrak{a} \arrow[r, "\xi"] & E
\end{tikzcd}$$

We know by Lemma 10.2 that λ, μ are defined over some finite extension of L, so after possibly enlarging it we may assume that also λ, μ are defined over L.

Step 4. We now reduce everything modulo $\mathfrak{q}$, to show that E_i is isomorphic to E^σ. Take a holomorphic differential ω on E which is rational over L, so that $\overline{\omega} \neq 0$. Then we have

$$\overline{[\mu \circ \lambda]}^*\overline{\omega} = \overline{[\alpha]}^*\overline{\omega} = \overline{\alpha\omega} = 0,$$

so that $\mu \circ \lambda$ reduces to an inseparable morphism. On the other hand, since $\mathfrak{x}$ is prime to p, we have that $\operatorname{Ker} \mu$ is of order $\operatorname{Nm} \mathfrak{x}$ and hence reduces to a separable morphism modulo $\mathfrak{q}$. As $\operatorname{Ker} \lambda$ is of order $\operatorname{Nm} \mathfrak{p} = p$, it follows that there is an isomorphism $\psi : \overline{E_i} \to \overline{E}^{(p)}$ such that the following triangle commutes:

$$\begin{tikzcd}
\overline{E} \arrow[r, "\overline{\lambda}"] \arrow[dr, "\mathrm{Frob}_E"'] & \overline{E_i} \arrow[d, "\psi"] \\
& \overline{E}^{(p)}
\end{tikzcd}$$

It follows that in fact $\overline{j(E_i)} = \overline{j(E)}^p = \overline{j(E)^\sigma}$, so that by our construction of L we must have that E and E_i are isomorphic.

Step 5. Using the fact that E^σ and E_i are isomorphic, we will now modify λ to obtain a *lift of Frobenius*, which is an isogeny κ whose reduction modulo $\mathfrak{q}$ is equal to the Frobenius morphism $\overline{E} \to \overline{E}^{(p)}$. We know that E^σ is isomorphic to E_i, so we may repeat the previous two steps with E^σ instead of E_i. This time, we obtain an automorphism

$$\psi : \overline{E}^{(p)} \to \overline{E}^{(p)}.$$

An easy calculation shows that the isomorphism $\operatorname{End}(E^\sigma) \simeq \mathcal{O}_K$ obtained from the normalised isomorphism $\operatorname{End}(E) \simeq \mathcal{O}_K$ is again normalised, so that we have the two equalities

$$\begin{cases} \overline{\lambda} \circ \overline{[a]} & = \quad \overline{[a]}^{(p)} \circ \overline{\lambda} \\ \mathrm{Frob}_E \circ \overline{[a]} & = \quad \overline{[a]}^{(p)} \circ \mathrm{Frob}_E \end{cases}$$

It follows that ψ commutes with $\overline{[a]}^{(p)}$, for all $a \in \mathcal{O}_K$. On the other hand, the endomorphism algebra $\operatorname{End}(\overline{E})$ was shown by Deuring to be an order in K, or a quaternion algebra. An element that commutes with all elements of the reduction of $\operatorname{End}(E)$ must therefore necessarily be in this reduction itself, so that ψ lifts to an isogeny $\widetilde{\psi} : E^\sigma \to E^\sigma$ which is necessarily an isomorphism. Setting $\kappa = \widetilde{\psi} \circ \lambda$ now yields a lift of Frobenius.

Step 6. We now show that the action of our lift of Frobenius κ on the m-torsion $E[m]$ (with m as chosen above) is via σ. Choose a point $t \in E[m]$, then we have $\overline{t^\sigma} = \mathrm{Frob}_p(\overline{t}) = \overline{\kappa(t)}$, so that the point $t^\sigma - \kappa(t)$ is an m-torsion point that reduces to 0 modulo $\mathfrak{q}$. On the other hand, since m is coprime to p, the reduction map $E[m] \to \overline{E}$ is an injection, so that we get the equality

$$t^\sigma = \kappa(t),$$

as required.

Step 7. To pass from $\mathfrak{p}$ to the idèle s corresponding to $\sigma\,|_{K^{\mathrm{ab}}}$ via the Artin map, factorise s as follows. Let $c \in \mathbf{A}_K^\times$ be an idèle with the property that $c_\mathfrak{p}$ is a uniformiser, and $c_v = 1$ for any $v \neq \mathfrak{p}$. As $K_{(m)}$ is contained in L, we see that

$$\varphi_{K_{(m)}/K}(s) = \varphi_{K_{(m)}/K}(c)$$

so that $c = sde$ for some $d \in K^\times$ and $e \in U_{(m)}$, where we use the notation from Appendix C. Since $\mathfrak{p}^{-1}\mathfrak{a} = c^{-1}\mathfrak{a} = d^{-1}s^{-1}\mathfrak{a}$, there is a unique isomorphism ξ_s such that the following diagram commutes:

$$\begin{tikzcd}
\mathbf{C} \arrow[r] \arrow[d, "\mathrm{Id}"'] & \mathbf{C}/\mathfrak{a} \arrow[r, "\xi"] \arrow[d] & E \arrow[d, "\kappa"] \\
\mathbf{C} \arrow[r] \arrow[d, "d"'] & \mathbf{C}/\mathfrak{p}^{-1}\mathfrak{a} \arrow[r, "\xi^*"] \arrow[d] & E^\sigma \arrow[d, "\mathrm{Id}"] \\
\mathbf{C} \arrow[r] & \mathbf{C}/s^{-1}\mathfrak{a} \arrow[r, "\xi_s"] & E^\sigma
\end{tikzcd}$$

Step 8. Finally, we show that ξ_s restricts to the torsion subgroup as required. Pick $u \in m^{-1}\mathfrak{a}$, and set $u_1 = u \pmod{\mathfrak{a}}$ and $u_2 = u \pmod{\mathfrak{p}^{-1}\mathfrak{a}}$. By the commutative diagram above, we get from step 6 that

$$\xi(u_1)^\sigma = \kappa(\xi(u_1)) = \xi^*(u_2) = \xi_s(du \bmod s^{-1}\mathfrak{a}).$$

The element du coincides with $s^{-1}u$ modulo $s^{-1}\mathfrak{a}$, as is easily seen by computing the local components:

$$\begin{cases} du = s_v^{-1}e_v^{-1}u \equiv s_v^{-1}u \pmod{s_v^{-1}\mathfrak{a}_v} \quad \text{for} \quad v \neq \mathfrak{p}, \\ du = s_\mathfrak{p}^{-1}c_\mathfrak{p}e_\mathfrak{p}^{-1}u \in s_\mathfrak{p}^{-1}(\mathfrak{p}\mathfrak{a})_\mathfrak{p}, \end{cases}$$

where we have used that $e_v - 1 \in m\mathcal{O}_v$ for all v, and c_v is a uniformiser when $v = \mathfrak{p}$ and 1 otherwise. It follows that

$$\xi(u \bmod \mathfrak{a})^\sigma = \xi_s(s^{-1}u \bmod s^{-1}\mathfrak{a}),$$

as required. Now let n be a multiple of m, then we may run through all of the above again, and obtain an isomorphism ξ'_s with the same property for n instead of m. We therefore must have

$$\xi'_s = [a]^\sigma \circ \xi_s,$$

for some a which must therefore be a unit in $\mathcal{O}_K$, and hence a root of unity of order dividing 6. By considering where both send an arbitrary element $v \in m^{-1}\mathfrak{a}/\mathfrak{a}$, we obtain from this that $av = v$, and hence a is a root of unity congruent to 1 modulo m. If m was at least 3 to start with, this implies that in fact $a = 1$. This implies that ξ_s satisfies the required property on n-torsion, for any multiple n of m, whence the conclusion. □

This wonderful theorem not only tells us that the Galois group acts through its abelianisation on the j-invariant of a CM elliptic curve, but it also gives us a very precise description of this action on the level of lattices. In the next section, we extract some concrete consequences for the explicit class field theory of imaginary quadratic extensions of K.

12 Abelian extensions of imaginary quadratic fields

The main theorem has far-reaching consequences, and in particular it allows us to recover a number of classical results on explicit class field theory due to Kronecker, Weber, Takagi, and Hasse. We will follow the exposition of Shimura [Shi70].

We continue to use the notation from last section. From the main theorem, it follows that $j(\mathfrak{a})^\sigma = j(s^{-1}\mathfrak{a})$, so that $j(\mathfrak{a})^\sigma$ only depends on the restriction of σ to K^{ab}. It follows that $j(\mathfrak{a}) \in K^{\mathrm{ab}}$, and the action of the element corresponding to s is via multiplication by s^{-1} on the lattice. We will make the nature of the numbers $j(\mathfrak{a})$ more precise in this section.

Coordinate functions. The main theorem explains how Galois acts on torsion points, so we would like to use these torsion points to generate abelian extensions of K. To talk about their coordinates in a meaningful way which is independent of the chosen model, we introduce the function h. Suppose $E : y^2 = 4x^3 + c_2x + c_3$, then we define

$$h_E((x,y)) = \begin{cases} c_2c_3/\Delta(E)\cdot x & \text{if} \quad j(E) \neq 0, 1728, \\ c_2^2/\Delta(E)\cdot x^2 & \text{if} \quad j(E) = 1728, \\ c_3/\Delta(E)\cdot x^3 & \text{if} \quad j(E) = 0. \end{cases}$$

An easy calculation shows that this function is independent of the chosen model, and that it is invariant under all automorphisms of E. In fact, its values capture precisely the orbits of points under automorphisms, in the sense that

$$h_E(t) = h_E(t') \iff t = \alpha t', \text{ some } \alpha \in \mathrm{Aut}(E).$$

These functions are often referred to as Weber functions, though we will reserve this name for the functions we have defined earlier. There is of course a relation between them, but as it stands they live in quite different universes so we will refrain from using this piece of nomenclature.

The coordinate function h_E defined above may be used to generate abelian extensions of K:

Theorem 12.1. *Let u be an element of $K/\mathfrak{a}$, and*

$$W = \left\{ s \in \mathbf{A}_K^\times \ : \ s\mathfrak{a} = \mathfrak{a}, su = u \right\}.$$

The field $K(j(E), h_E(\xi(u)))$ is the subfield of K^{ab} corresponding to the subgroup $K^\times W$ of $\mathbf{A}_K^\times$.

Proof. The group $W \subset \mathbf{A}_K^\times$ is open, as it contains the open subgroup $U_{(m)}$ whenever $mu \in \mathfrak{a}$. It clearly also contains the group of infinite idèles $\mathbf{A}_K^{f,\times} = \mathbf{C}^\times$. The open subgroup $K^\times W$ corresponds via class field theory to an abelian extension F/K. Let $\sigma \in \mathrm{Aut}(\mathbf{C}/K)$, and take an idèle s so that $\sigma\mid_{K^{\mathrm{ab}}} = \varphi(s)$. The main theorem provides us with an isomorphism $\xi_s : \mathbf{C}/s^{-1}\mathfrak{a} \to E^\sigma$.

Suppose first that σ restricts to the identity map on F, then we can assume that $s \in W$. As $s\mathfrak{a} = \mathfrak{a}$ we have that $j(E)^\sigma = j(E)$, so that we can find an isomorphism $\psi : E \to E^\sigma$ satisfying $\psi \circ \xi_s = \xi$. As the coordinate functions h_E are invariant under automorphisms, we get

$$h_E(\psi \circ \xi(u)^\sigma) = h_{E^\sigma}(\xi(u)^\sigma) = h_E(\xi(u))^\sigma.$$

In addition, we have that $\psi \circ \xi(u)^\sigma = \psi \circ \xi_s(s^{-1}u) = \xi(u)$ so that we obtain $h_E(\xi(u))^\sigma = h_E(\xi(u))$. This shows that σ acts trivially on $K(j(E), h_E(\xi(u))$ as well.

Conversely, suppose σ restricts to the identity map on $K(j(E), \xi(u))$. Then there is an isomorphism $\psi : E^\sigma \to E$, which means there is an element $\mu \in K^\times$ so that $\mu s^{-1}\mathfrak{a} = \mathfrak{a}$. By choosing a suitable isomorphism ψ, we may assume the following diagram commutes:

$$\begin{tikzcd}
\mathbf{C} \arrow[r] \arrow[d, "\mu"'] & \mathbf{C}/s^{-1}\mathfrak{a} \arrow[r, "\xi_s"] \arrow[d] & E^\sigma \arrow[d, "\psi"] \\
\mathbf{C} \arrow[r] & \mathbf{C}/\mathfrak{a} \arrow[r, "\xi"] & E
\end{tikzcd}$$

Because σ acts trivially on $h_E(\xi(u))$, we obtain

$$h_E(\psi \circ \xi(u)^\sigma) = h_{E^\sigma}(\xi(u)^\sigma) = h_E(\xi(u))^\sigma = h_E(\xi(u)),$$

which implies the existence of $\zeta \in K^\times$ such that

$$\zeta\mathfrak{a} = \mathfrak{a}, \quad \text{and} \quad [\zeta] \circ \psi \circ \xi(u)^\sigma = \xi(u).$$

We have that $\psi \circ \xi(u)^\sigma = \psi \circ \xi_s(s^{-1}u) = \xi(\mu s^{-1}u)$, so if we put $s' = \zeta\mu s^{-1}$ we check that $s'\mathfrak{a} = \mathfrak{a}$ and $s'u = u$, so that $s' \in W$ and hence $s \in K^\times W$. This implies that σ acts trivially on F. □

This theorem implies that in fact, we may generate all of K^{ab} using j-invariants of CM elliptic curves, as well as values of h_E on torsion points.

Corollary 12.1. *K^{ab} is generated over K by $j(E)$ and the values $h_E(t)$ for all torsion points t on E.*

Proof. If $s \in \mathbf{A}_K^\times$ that is contained in all the subgroups W as in the previous theorem, we must have $(s-1)u \in \mathfrak{a}$ for all $u \in K$. This implies that s must have local factor 1 at all primes, whence $s \in \mathbf{A}_K^{f,\times} = \mathbf{C}^\times$. The intersection of the groups $K^\times W$ is therefore precisely the kernel of the Artin map $\varphi_K : \mathbf{A}_K^\times \to \text{Gal}(K^{\text{ab}}/K)$. The statement now follows from the previous theorem. □

Since any finite abelian extension of K is contained in a ray class field for some modulus, we wonder which particular combination of the torsion coordinates we need to generate some ray class field. The following corollary does this for those corresponding to principal ideals, see the exercises for the general case.

Corollary 12.2. *The field $K(j(E), h_E(E[m]))$ is the ray class field $K_{(m)}$.*

Proof. By the above theorem, the field $K(j(E), h_E(E[m]))$ corresponds to the subgroup $K^\times U$, where U is the group of idèles s such that

$$s\mathfrak{a} = \mathfrak{a}, \quad \text{and} \quad su = u, \ {}^\forall u \in m^{-1}\mathfrak{a}.$$

This means that s is an idèle which is a unit at all finite places, and at those places dividing m it must be $1 \pmod{m}$. This is precisely the definition of the ray class group $U_{(m)}$. □

We now come to a more precise description of the individual algebraic numbers $j(\mathfrak{a})$. The most important result is:

Corollary 12.3. *Let $\mathfrak{a}$ be a proper ideal of an order $\mathcal{O}$ in K, then we have*

- *$K(j(\mathfrak{a}))$ is the ring class field of $\mathcal{O}$ over K,*
- *The map $\mathrm{Gal}(K(j(\mathfrak{a}))/K) \to \mathrm{Cl}\ \mathcal{O} : \sigma \mapsto \mathfrak{b}$, where $j(\mathfrak{a})^\sigma = j(\mathfrak{b}^{-1}\mathfrak{a})$, is an isomorphism,*
- *We have $\deg(K(j(\mathfrak{a}))/K) = \deg(\mathbf{Q}(j(\mathfrak{a}))/\mathbf{Q})$,*

Proof. By setting $u = 0$ in the above theorem, we see that $K(j(\mathfrak{a}))$ corresponds to the subgroup

$$W = \mathbf{A}_K^{f,\times} \prod_{\mathfrak{p}} \mathcal{O}_{\mathfrak{p}}^\times,$$

which is precisely the subgroup used to define the ring class field of $\mathcal{O}$. A $\mathbf{Z}$-lattice is a proper $\mathcal{O}$-ideal if and only if it is of the form $s\mathcal{O}$ for some $s \in \mathbf{A}_K^\times$ (see for instance Shimura), so that $s \mapsto s\mathcal{O}$ gives an isomorphism

$$\mathbf{A}_K^\times / K^\times W \xrightarrow{\sim} \mathrm{Cl}\ \mathcal{O}.$$

This proves the first two statements. Let E be an elliptic curve isomorphic to $\mathbf{C}/\mathfrak{a}$, and $\sigma \in \mathrm{Aut}(\mathbf{C})$. Then $\mathrm{End}(E^\sigma) \simeq \mathrm{End}(E)$, so that $j(\mathfrak{a})^\sigma = j(\mathfrak{b})$ for some proper $\mathcal{O}$-ideal $\mathfrak{b}$. This means that $[\mathbf{Q}(j(\mathfrak{a})) : \mathbf{Q}] \leq [K(j(\mathfrak{a})) : K]$, and so we must have equality as required. □

13 Three proofs of integrality of j

Now that we know that the j-invariant of an elliptic curve with complex multiplication is an algebraic number, we see that we can only consider this to be a satisfactory method to explicitly construct abelian extension of an imaginary quadratic field K if we can effectively determine the algebraic values $j(\tau)$. As a first step, we will show that these values are integral, in three different ways.

13.1 The modular polynomial.

First, let us give a complex analytic proof of the fact that $j(\tau)$ is integral whenever E_τ has CM by an order in K. The proof uses modular polynomials, which made their first appearance last term, when Prof. Darmon used them to show that the modular curves $X_0(N)$ are algebraic curves, defined over $\mathbf{Q}$. Recall that the modular polynomial $\Phi_N(X, Y) \in \mathbf{Z}[X, Y]$ is an irreducible symmetrix polynomial which is monic in both variables, of the same degree. It is further characterised by the property

$$\Phi_N(j(z), j(Nz)) = 0.$$

This polynomial is the main ingredient in the proof of the following result:

Theorem 13.1. *Let $E_{\mathbf{C}}$ be a CM elliptic curve. Then $j(E)$ is an algebraic integer.*

Proof. Consider first the polynomial $\Delta_N(X) = \Phi_N(X,X)$, which has coefficients in $\mathbf{Z}$. Consider the meromorphic function $\Delta(j(q))$, which by construction is equal to the product

$$\prod_{\alpha \in \mathcal{S}_N} (j(q) - j^{\alpha}(q)), \qquad \text{where } \mathcal{S}_N = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in M_2(\mathbf{Z}) \ : ad = N, d > 0, 0 \leq b < d \right\}.$$

An elementary calculation on q-expansions reveals that every factor is of the form

$$\frac{1}{q} - \frac{\zeta_N^{-ab}}{q^{a/d}} + \text{holo.}$$

If N is not a perfect square, this shows that the leading coefficient of each factor is always a root of unity. Therefore, the same is true for $\Delta(j(q))$, and as $\Delta_N \in \mathbf{Z}[x]$ has integer coefficients, this shows that its leading term must be ± 1.

Let E be as in the statement, then $\mathrm{End}(E)$ is an order in an imaginary quadratic field K. As $j(\alpha\tau)$ is integral over $\mathbf{Z}[j(\tau)]$ for any integral matrix α with positive determinant, we may assume without loss of generality that $\mathrm{End}(E) = \mathcal{O}_K$, the maximal order. Now choose an element $\lambda \in \mathcal{O}_K$ such that $N = |\mathrm{Nm}_{K/\mathbf{Q}}\lambda|$ is not a perfect square. This implies that there exists an isogeny

$$[\lambda] : E \to E, \quad \text{such that} \quad \deg[\lambda] = N,$$

and hence $j(E)$ is a root of $\Delta_N[X] \in \mathbf{Z}[X]$, which we showed has leading coefficient ± 1. □

As the modular polynomials are computable in practice, this gives us a way to find explicit minimal polynomials for the numbers $j(E)$. The heights of these polynomials tend to be outrageous, and in general this yields a highly impractical way of determining minimal polynomials of ring class fields.

13.2 The criterion of Néron–Ogg–Shafarevich.

We now present an argument of Serre–Tate [ST68] to show that the j-invariant of a CM elliptic curve is integral. It uses the criterion of Néron–Ogg–Shafarevich:

Theorem 13.2. *Let A be an abelian variety over a local field K, and let l be a prime different from the residue characteristic of K. Then A has good reduction over K if and only if $T_l A$ is unramified at p.*

We will first show how this implies an important foundational result about the reduction of CM elliptic curves:

Theorem 13.3. *Let $E_{\mathbf{C}}$ be a CM elliptic curve. Then E is defined over a number field $L/\mathbf{Q}$, and has potentially good reduction at every prime of L.*

Proof. By the main theorem, we know that E is defined over a number field L. Fix a finite prime $\mathfrak{p}$ of L, lying above a prime p of $\mathbf{Q}$. Choose a prime $l \neq 2, p$, and consider the action of the inertia group $I_\mathfrak{p} \trianglelefteq G_\mathfrak{p} = \mathrm{Gal}(\overline{L_\mathfrak{p}}, L_\mathfrak{p})$ on the l-adic Tate module $T_l E$. By the main theorem of complex multiplication, we know that $I_\mathfrak{p}$ acts through an abelian quotient. By local class field theory, we have

$$I_\mathfrak{p}^{\mathrm{ab}} \simeq \mathcal{O}_\mathfrak{p}^\times \simeq (1 + \mathfrak{p}\mathcal{O}_\mathfrak{p}) \times (\mathcal{O}_\mathfrak{p}/\mathfrak{m})^\times.$$

Note that $(1 + \mathfrak{p}\mathcal{O}_\mathfrak{p})$ is a pro-p group. This leaves us with a morphism $\psi : I_\mathfrak{p}^{\mathrm{ab}} \to \mathrm{GL}_2(\mathbf{Z}_l)$, and we have the exact sequence coming from mod l reduction:

$$1 \to 1 + lM_2(\mathbf{Z}_l) \to \mathrm{GL}_2(\mathbf{Z}_l) \to \mathrm{GL}_2(\mathbf{F}_l) \to 1.$$

As $1 + lM_2(\mathbf{Z}_l)$ is a pro-l group, it cannot intersect the pro-p group $\psi(1 + \mathfrak{p}\mathcal{O}_\mathfrak{p})$, so that the image of ψ must be finite. The fixed field $F_\mathfrak{q}$ of Ker ψ acting on the maximal totally ramified abelian extension of $L_\mathfrak{q}$ is therefore a finite extension of $L_\mathfrak{p}$, and over $F_\mathfrak{q}$ the inertia subgroup acts trivially on the l-adic Tate module, so that E has good reduction by the criterion of Néron–Ogg–Shafarevich. □

This is a beautiful and important auxiliary result on CM elliptic curves. As a consequence, we obtain a second proof of the integrality of $j(E)$.

Theorem 13.4. *Let $E_\mathbf{C}$ be a CM elliptic curve. Then $j(E)$ is an algebraic integer.*

Proof. If the algebraic number $j(E)$ were not integral at a prime $\mathfrak{p}$ of $K(j(E))$, then over any finite extension the curve E would have bad reduction at all primes above $\mathfrak{p}$, contradiction. □

13.3 Tate uniformisation.

It is also possible to use the p-adic uniformisation of Tate curves to show that if $j(E)$ is not integral at a prime $\mathfrak{p}$, the Galois action on the p-adic Tate module contains quasi-unipotent elements, which forces the endomorphism ring to be trivial. We will discuss the main steps in this proof, as it involves some beautiful results on p-adic geometry.

Recall that the Tate curve $\mathrm{Tate}(q)$ is the elliptic curve $y^2 + xy = x^3 + a_4 x + a_6$, where

$$a_4 = -\sum_{n \geq 1} \frac{5n^3 q^n}{1 - q^n}, \qquad a_6 = -\sum_{n \geq 1} \frac{1}{12} \cdot \frac{(7n^5 + 5n^3) q^n}{1 - q^n}.$$

This defines an elliptic curve over $\mathbf{Z}((q))$ whose discriminant and j-invariant give us the q-expansions of the modular forms $\Delta(q)$ and $j(q)$. For $\tau \in \mathfrak{H}$ the series defining $\mathrm{Tate}(q)$ converges at $q = e^{2\pi i \tau}$ and defines an elliptic curve isomorphic to $\mathbf{C}/(\mathbf{Z} + \mathbf{Z}\tau)$. Define the series

$$\begin{cases} X(u,q) &= \frac{u}{(1-u)^2} + \sum_{d \geq 1} \left(\sum_{m|d} m(u^m + u^{-m} - 2) \right) q^d, \\ Y(u,q) &= \frac{u^2}{(1-u)^3} + \sum_{d \geq 1} \left(\sum_{m|d} \frac{m(m-1)}{2} u^m - \frac{m(m+1)}{2} u^{-m} + m \right) q^d. \end{cases}$$

For any $q \in \overline{\mathbf{Q}}_p$ such that $|q| < 1$, the power series $X(u,q)$ and $Y(u,q)$ converge for all $u \in \overline{\mathbf{Q}}_p^{\times} \backslash q^{\mathbf{Z}}$, and they define a surjective, Galois equivariant homomorphism

$$\pi : \overline{\mathbf{Q}}_p^{\times} \to \mathrm{Tate}(q)(\overline{\mathbf{Q}}_p) : u \mapsto (X(u,q), Y(u,q)),$$

with kernel $q^{\mathbf{Z}}$. Now suppose we have any finite extension $K/\mathbf{Q}_p$, as well as an elliptic curve

$$E_K : y^2 = x^3 + ax + b.$$

Note that any elliptic curve has a Weierstrass equation of this form. It turns out that there is a criterion for E_K to be isomorphic over $\overline{K}$ to the Tate curve $\mathrm{Tate}(q)$ for some value of q.

Theorem 13.5 (Tate). *Suppose $|j(E)| > 1$, then there is a unique $q \in \overline{K}^{\times}$ with $v_p(q) > 0$ such that there is a $\overline{K}^{\times}$-isomorphism*

$$\psi : E \to \mathrm{Tate}(q).$$

We have $q \in K$, and we may choose ψ to be defined over K if and only if $2a/b \in (K^{\times})^2$.

This result, combined with the existence of the power series X, Y above, is usually referred to as the *p-adic uniformisation of Tate curves* or simply *Tate uniformisation.* This is justified by the fact that X, Y define a genuine isomorphism of *rigid analytic spaces*

$$\mathbf{G}_m/q^{\mathbf{Z}} \xrightarrow{\sim} E.$$

This has been generalised to other types of curves by Mumford, who showed that a similar p-adic version of uniformisation for Riemann surfaces holds for curves whose reduction is totally degenerate, meaning all its irreducible components are isomorphic to $\mathbf{P}^1$, crossing transversally.

Rather than proving the statements on Tate uniformisation, we will show how to put it in action to prove, once more, that j-invariants of CM elliptic curves are algebraic integers.

Theorem 13.6. *Let $E_{\mathbf{C}}$ be a CM elliptic curve. Then $j(E)$ is an algebraic integer.*

Proof. Suppose that the algebraic number $j(E)$ is non-integral at some prime $\mathfrak{p}$ of $L = K(j(E))$. Choose a model of E over $L_{\mathfrak{p}}$, then as $|j(E)| > 1$ there exists a $q \in L_{\mathfrak{p}}$ such that

$$\mathbf{G}_m/q^{\mathbf{Z}} \xrightarrow{\sim} E,$$

and moreover this isomorphism is Galois equivariant. Now choose l large enough so that

$$l \nmid \mathrm{ord}_{\mathfrak{p}}(j(E)),$$

and enlarge $L_{\mathfrak{p}}$ to contain an l-th root of unity ζ_l. Note that after this enlargement we still have $l \nmid \mathrm{ord}_{\mathfrak{p}}(j(E))$. Consider the extension $L_{\mathfrak{p}}(\sqrt[l]{q})/L_{\mathfrak{p}}$, which is a totally ramified Galois extension of degree l. As it is a Kummer extension, we see that the inertia group contains an element σ_l such that

$$\sqrt[l]{q}^{\sigma_l} = \zeta_l \sqrt[l]{q}.$$

The l-torsion of $\mathbf{G}_m/q^{\mathbf{Z}}$ is generated by ζ_l and $\sqrt[l]{q}$. By the p-adic uniformisation of E, we see that as $\mathrm{Gal}(\overline{L}_{\mathfrak{p}}/L_{\mathfrak{p}})$-modules, we have a short exact sequence

$$0 \to \underline{\mathbf{Z}/l\mathbf{Z}} \to E[l] \to \langle \sqrt[l]{q} \rangle \to 0.$$

This shows that for any prime l that does not divide $\mathrm{ord}_{\mathfrak{p}}(j(E))$, there is an element $\sigma_l \in \mathrm{Gal}(\overline{L}/L)$ which acts on $E[l]$ as the matrix

$$\sigma_l = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

with respect to a suitable basis.

Now choose a non-zero endomorphism $\lambda \in \mathrm{End}(E)$. After possibly enlarging L, we may assume that λ is defined over L. Let l be a prime which is large enough, so that we have an element σ_l as constructed above. As λ is defined over L, it commutes with the Galois action, so that by considering its action on $T_l E$ we get, for an appropriate basis, the following equation for its entries $a, b, c, d \in \mathbf{Z}_l$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \pmod{l}.$$

This equation yields $a \equiv d \pmod{l}$ and $c \equiv 0 \pmod{l}$. Now let $m = \deg(1+\lambda) - \deg(\lambda) - 1$, which can be calculated by taking the determinant of the corresponding matrices on $T_l E$ to satisfy $m \equiv 2a \pmod{l}$. This implies that

$$\deg([m] - 2\lambda) \equiv 0 \pmod{l}.$$

As we have this congruence for infinitely many primes l, we see that in fact, $2\lambda = [m]$. This implies that λ itself must be multiplication by some integer, and hence E cannot be CM. □

14 The class number one problem

In this section, we will discuss Heegner's proof of the class number one problem. More precisely, we prove the following result:

Theorem 14.1. *There are precisely* 9 *imaginary quadratic fields of class number* 1*:*

$$\mathbf{Q}(\sqrt{-d}), \qquad \textit{where } d = 1, 2, 3, 7, 11, 19, 43, 67, 163.$$

The proof will be given in the following paragraphs. We start by dealing with the case $\Delta_K \equiv 0 \pmod{4}$, using only reduction theory of quadratic forms. The case $\Delta_K \equiv 1 \pmod{4}$ is the hardest, and before we deal with it, we compute some special values of the j-function and γ_2-function. Then, we use the relations between Weber functions to show how any quadratic imaginary field with class number one gives rise to an integer solution to a certain Diophantine equation, which we may explicitly solve. All possible solutions will be accounted for by the tables of special values we computed, concluding the proof.

14.1 Quadratic forms.

We can already prove a weaker statement, using only reduction theory of definite quadratic forms! More precisely, we have the following theorem, which is due to Landau:

Theorem 14.2. *Let n be a positive integer, then $h(-4n) = 1$ if and only if $n = 1, 2, 3, 4, 7$.*

Proof. If $n = 1, 2, 3, 4, 7$ then it is easy to check that the class number is 1, by simply listing the reduced forms as explained in Chapter I. Conversely, assume that $h(-4n) = 1$, so we know that there is precisely one reduced form which therefore must be $\langle 1, 0, n\rangle$. If n is not a prime power, then we may write $n = ac$ with $(a, c) = 1$ and $1 < a < c$, and we find a second reduced form $\langle a, 0, c\rangle$, which is a contradiction. Therefore n must be a prime power.

If $n = 2^r$, then we cannot have $r \geq 4$ or else $\langle 4, 4, 2^{r-2} + 1\rangle$ would be reduced. This leaves us with 4 possible cases, which we can check by hand to see that $n = 1, 2, 4$.

If $n = p^r$ for p odd, then $n + 1$ must also be a prime power or else $n + 1 = ac$ with $1 < a < c$ and $(a, c) = 1$ which would produce the reduced form $\langle a, 2, c\rangle$. As p is odd, we must have $n + 1 = 2^s$. We see that $s \leq 5$, as otherwise $\langle 8, 6, 2^{s-3} + 1\rangle$ would be a reduced form. This leaves us with finitely many cases, and we check by hand that only $n = 1, 3, 7$ gives class number one. □

The case $\Delta_K \equiv 1 \pmod 4$ lies much deeper, and we will use CM theory to solve it. More precisely, CM theory tells us about the integrality of special values of certain modular functions. The relations between Weber functions that we have established then give rise to Diophantine equations that we will solve explicitly.

14.2 Integer j-invariants.

First, we will gather some numerical data concerning the nine fields appearing in the above theorem. This data will be used in the next section.

We start with the j-invariant. Let K be an imaginary quadratic field with class number one. We know that $j(\mathcal{O}_K) \in K$. In fact, for any $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, we have that $j(\mathcal{O}_K)^\sigma$ is the j-invariant of an elliptic curve with CM by $\mathcal{O}_K$, so that it must be $j(\mathcal{O}_K)$, and hence $j(\mathcal{O}_K) \in \mathbf{Q}$. By the integrality of the j-invariant, we even have $j(\mathcal{O}_K) \in \mathbf{Z}$. This allows us to compute guesses for j-invariants numerically, and use these to find Weierstrass equations for elliptic curves with CM by $\mathcal{O}_K$. As an example, let us compute the value

$$j\left(\tau\right), \qquad \text{where } \tau = \frac{1 + \sqrt{-163}}{2},$$

by explicitly evaluating the q-expansion for the j-function at $q = e^{2\pi i\tau} = -e^{-\pi\sqrt{163}} \approx -3.809 \cdot 10^{-18}$, which is an extremely small number. This shows that the main contributions to $j(\tau)$ come from the terms q^{-1} and 744, and we compute numerically that

$$q^{-1} \approx -262537412640768743.999999999999250072597$$

from which we guess that in fact $j(\tau)$ converges to the rational integer $-2^{18} \cdot 3^3 \cdot 5^3 \cdot 23^3 \cdot 29^3$.

For all the fields appearing in the statement of the class number one problem, we may likewise numerically compute the j-invariant. In fact, for future reference we will also tabulate Weierstrass equations over $\mathbf{Q}$ for the unique isomorphism class of elliptic curves over that have CM by the corresponding maximal order. We note that once these Weierstrass equations are found, we can in theory verify that they do indeed have CM by $\mathcal{O}_K$, making all computations rigorous.

Field	$E_{\mathbf{Q}}$ with CM by maximal order	$j(E)$
$\mathbf{Q}(\sqrt{-1})$	$y^2 = x^3 + x$	$2^6 \cdot 3^3$
$\mathbf{Q}(\sqrt{-2})$	$y^2 = x^3 + x$	$2^6 \cdot 5^3$
$\mathbf{Q}(\sqrt{-3})$	$y^2 + xy = x^3 - x^2 - 2x - 1$	0
$\mathbf{Q}(\sqrt{-7})$	$y^2 = x^3 + 4x^2 + 2x$	$-3^3 \cdot 5^3$
$\mathbf{Q}(\sqrt{-11})$	$y^2 + y = x^3 - x^2 - 7x + 10$	-2^{15}
$\mathbf{Q}(\sqrt{-19})$	$y^2 + y = x^3 - 38x + 90$	$-2^{15} \cdot 3^3$
$\mathbf{Q}(\sqrt{-43})$	$y^2 + y = x^3 - 860x + 9707$	$-2^{18} \cdot 3^3 \cdot 5^3$
$\mathbf{Q}(\sqrt{-67})$	$y^2 + y = x^3 - 7370x + 243528$	$-2^{15} \cdot 3^3 \cdot 5^3 \cdot 11^3$
$\mathbf{Q}(\sqrt{-163})$	$y^2 + y = x^3 - 2174420x + 1234136692$	$-2^{18} \cdot 3^3 \cdot 5^3 \cdot 23^3 \cdot 29^3$

Notice a curious fact: All these j-invariants are perfect cubes! This shows that in fact also the function γ_2 takes values in $\mathbf{Z}$. As we will see now, this is not a coincidence, as often ray class fields are actually generated by the values of γ_2, and we will need one instance of this in what follows:

Theorem 14.3. *Let $\Delta < 0$ be a discriminant not divisible by 3, and set*

$$\tau_0 = \begin{cases} \sqrt{-m} & \text{if} \quad \Delta = -4m, \\ \frac{3+\sqrt{-m}}{2} & \text{if} \quad \Delta = -m \equiv 1 \pmod 4). \end{cases}$$

Then $\gamma_2(\tau_0)$ is an algebraic integer, and $K(\gamma_2(\tau_0))$ is the ring class field of $\mathbf{Z}[\tau_0]$. Moreover, we have $\mathbf{Q}(\gamma_2(\tau_0)) = \mathbf{Q}(j(\tau_0))$.

Proof. Todo. □

It is precisely the function γ_2, through its connection with the Weber functions, that will play a central role in the proof of the class number one problem. Let us compute some of its values. In contrast to the cavalier approach we took to the approximation of the j-invariant above, we will now use the connection with the Weber functions, which have explicit product expansions that we can estimate, to ensure that we know how many terms to compute in order to know with absolute certainty that the integers computed numerically are indeed correct.

Set $\tau_0 = \frac{-3+\sqrt{-p}}{2}$. We will now describe how to compute $\gamma_2(\tau_0)$, for $p = 3, 11, 19, 43, 67, 163$. Recall that we have the relations

$$\gamma_2(\tau_0) = \mathfrak{f}_2(\tau_0)^{16} + \frac{16}{\mathfrak{f}_2(\tau_0)^8}, \quad \mathfrak{f}_2(\tau_0) = \frac{\sqrt{2}}{\mathfrak{f}_1(2\tau_0)}.$$

The transformation laws for the Weber functions also show that $\mathfrak{f}_1(2\tau_0) = \zeta_{48}^{-3}\mathfrak{f}(\sqrt{-p})$, from which we finally deduce that

$$\gamma_2(\tau_0) = \frac{256}{\mathfrak{f}(\sqrt{-p})^{16}} - \mathfrak{f}(\sqrt{-p})^8.$$

Therefore, it suffices to compute $\mathfrak{f}(\sqrt{-p})$. Using the product expansion of $\mathfrak{f}$, we will estimate this value precisely. Set $q = e^{-2\pi\sqrt{-p}}$. Using the elementary inequality $1 + x < e^x$, whenever $x > 0$, we get

$$\prod_{n=1}^{\infty}(1 - q^{n-1/2}) < \prod_{n=1}^{\infty} e^{q^{n-1/2}} = e^{\sqrt{q}/(1-q)}.$$

The exponent may be estimated further by $\sqrt{q}/(1-q) \leq \sqrt{q}/(1-e^{-2\pi}) < 1.002\sqrt{q}$, which by the product formula for the Weber function $\mathfrak{f}$ yields the estimate

$$q^{-1/48} < \mathfrak{f}(\sqrt{-p}) < q^{-1/48}e^{1.002\sqrt{q}}.$$

This gives us the following bound on $\gamma_2(\tau_0)$:

$$256q^{1/3} - q^{-1/6}e^{8.016\sqrt{q}} < \gamma_2(\tau_0) < 256q^{1/3}e^{-16.032\sqrt{q}} - q^{-1/6}$$

As we have $q < e^{-2\pi}$, this estimate is sharp enough to obtain

$$\gamma_2(\tau_0) = \left[-q^{-1/6} + 256q^{1/3}\right],$$

where the square brackets denote the function that rounds to the nearest integer. This results in an effective way of computing all the relevant values of $\gamma_s(\tau_0)$:

Field	τ_0	$\gamma_2(\tau_0)$
$\mathbf{Q}(\sqrt{-3})$	$\frac{1+\sqrt{-3}}{2}$	0
$\mathbf{Q}(\sqrt{-11})$	$\frac{3+\sqrt{-11}}{2}$	-2^5
$\mathbf{Q}(\sqrt{-19})$	$\frac{3+\sqrt{-19}}{2}$	$-2^5 \cdot 3$
$\mathbf{Q}(\sqrt{-43})$	$\frac{3+\sqrt{-43}}{2}$	$-2^6 \cdot 3 \cdot 5$
$\mathbf{Q}(\sqrt{-67})$	$\frac{3+\sqrt{-67}}{2}$	$-2^5 \cdot 3 \cdot 5 \cdot 11$
$\mathbf{Q}(\sqrt{-163})$	$\frac{3+\sqrt{-163}}{2}$	$-2^6 \cdot 3 \cdot 5 \cdot 23 \cdot 29$

We now turn to the class number one problem, which we will solve by showing that the numbers $\gamma_2(\tau_0)$ for any p corresponding to a quadratic imaginary field of class number one, give us an integer solution to a certain Diophantine equation. As we will see, all solutions correspond to an entry in the above table.

We will need one additional ingredient. Given that the values of the Weber function were used to compute the numbers $\gamma_2(\tau_0)$ explicitly, we wonder whether those values themselves generate ring class fields. The following theorem shows that this is true under certain conditions.

Theorem 14.4. *Let m be a positive integer such that $3 \nmid m$ and $m \equiv 3 \pmod 4$. Then $\alpha = \mathfrak{f}(\sqrt{-m})^2$ is an algebraic integer and $K(\alpha)$ is the ring class field of $\mathbf{Z}[\sqrt{-m}]$.*

Proof. Todo. □

14.3 The class number one problem: proof.

We now discuss the arguments of Heegner, as presented by Stark and Birch, to prove the class number one problem.

Theorem 14.5. *There are precisely 9 imaginary quadratic fields of class number 1:*

$$\mathbf{Q}(\sqrt{-d}), \qquad \textit{where } d = 1, 2, 3, 7, 11, 19, 43, 67, 163.$$

Proof. The proof is quite long, so we will break it up into several steps. Conceptually, the most important steps are the reduction to the Diophantine problem considered below, using the identities between Weber functions we established.

Step 1. First, we reduce the problem to a particular case where the ring class field is a cubic extension of $K = \mathbf{Q}(\sqrt{-d})$. Let Δ be the discriminant of K, and suppose that $h(\Delta) = 1$. We have dealt with the case where Δ is even in Theorem 14.2, so we may assume that $\Delta \equiv 1 \pmod 4$. By genus theory, we must have $\Delta = -p$ for some prime p. If $p \equiv 7 \pmod 8$, the class number formula (see exercises) tells us

$$h(-4p) = 2h(-p)\left(1 - \frac{1}{2}\left(\frac{-p}{2}\right)\right) = 1,$$

so that we must have $p = 7$ by Theorem 14.2. We therefore are reduced to dealing with the case $p \equiv 3 \pmod 8$, and may assume $p \neq 3$. Again, by the class number formula, we obtain

$$h(-4p) = 3.$$

It follows that the ring class field $K(j(\sqrt{-p}))$ of $\mathbf{Z}[\sqrt{-p}]$ is a degree 3 extension, and we easily deduce that $\mathbf{Q}(j(\sqrt{-p}))$ is real. Since $\mathfrak{f}(\sqrt{-p})$ is real, it follows from the results in the previous section that

$$[\mathbf{Q}(\mathfrak{f}(\sqrt{-p})^2) : \mathbf{Q}] = 3.$$

Step 2. The main step is the construction of two algebraic numbers α and α^4. Start by defining

$$\tau_0 = \frac{3 + \sqrt{-p}}{2},$$

and set $\alpha = \zeta_8 \mathfrak{f}_2(\tau_0)^2$. From the functional equations of the Weber functions, we get

$$\mathfrak{f}_1(2\tau_0) = \mathfrak{f}_1(3 + \sqrt{-p}) = \zeta_{48}^{-3} \mathfrak{f}(\sqrt{-p}),$$

so that it follows from the identity $\mathfrak{f}_1(2\tau_0)\mathfrak{f}_2(\tau_0) = \sqrt{2}$ that

$$\alpha = \frac{2}{\mathfrak{f}(\sqrt{-p})^2}.$$

It follows that

$$\mathbf{Q}(\mathfrak{f}(\sqrt{-p})^2) = \mathbf{Q}(\alpha) = \mathbf{Q}(\alpha^4).$$

Step 3. Now we will see that this puts such severe restrictions on the minimal polynomials of α and α^4 that we will be able to find out all possibilities for τ_0 explicitly. By Corollary 9.1, we know that α^4 satisfies the polynomial

$$x^3 - \gamma_2(\tau_0)x - 16 = 0.$$

This is a polynomial in $\mathbf{Z}[x]$, and as α^4 generates a degree 3 extension of $\mathbf{Q}$ it must be its minimal polynomial. This means that also α must be an algebraic integer, and hence it satisfies a polynomial

$$x^3 + ax^2 + bx + c = 0,$$

with $a, b, c \in \mathbf{Z}$. By moving the even degree terms to the other side, and squaring both sides, we see that α satisfies

$$x^6 + (2b - a^2)x^4 + (b^2 - 2ac)x^2 - c^2 = 0.$$

Moving the terms with degree divisible by 4 to the other side, and squaring both sides, we finally obtain that α must satisfy

$$x^{12} + (2f - e^2)x^8 + (f^2 - 2eg)x^4 - g^2 = 0,$$

where $e = (2b - a^2)$, $f = (b^2 - 2ac)$, and $g = -c^2$. As a polynomial in x^4, this must be equal to the minimal polynomial of α^4, whence we obtain

$$\begin{cases} 2f - e^2 & = & 0, \\ f^2 - 2eg & = & -\gamma_2(\tau_0), \\ g^2 & = & 16. \end{cases}$$

From the last equation, we get $c = \pm 2$. By changing the sign of α if necessary, we may assume $c = 2$. The first equation of the above system now becomes

$$2(b^2 - 4a) = (2b - a^2)^2.$$

It is elementary to check that this implies that a and b must both be even. This means that we obtain an integer solution to the equation

$$2X(X^3 + 1) = Y^2, \tag{II.2}$$

by setting $X = -a/2$ and $Y = (b - a^2)/2$.

Step 4. We will now describe how to find all solutions to (II.2). Note that for any solution (X, Y), the numbers X and $X^3 + 1$ are coprime, and hence it gives rise to a solution of one of the following equations:

- $x^3 + 1 = y^2$,
- $x^3 + 1 = -y^2$,
- $x^3 + 1 = 2y^2$,
- $x^3 + 1 = -2y^2$.

Each of these cases may be solved separately. This is quite a lengthy process, so we will just treat one case as an example. For full details, see [Cox89]. We also note that the first equation is the most difficult one, and requires some clever algebraic manipulations.

The equation $x^3 + 1 = -y^2$. In the ring $\mathbf{Z}[i]$, which is a unique factorisation domain, we may factorise the equation as

$$(y + i)(y - i) = -x^3.$$

It is easy to show that $(y + i)$ and $(y - i)$ are coprime, so they must both be cubes. Setting $y + i = (a + bi)^3$ for $a, b \in \mathbf{Z}$ gives us $b(3a^2 - b^2) = 1$ from which we deduce that $(x, y) = (-1, 0)$ is the only solution.

The equations $x^3 + 1 = 2y^2$ and $x^3 + 1 = -2y^2$ may be dealt with similarly, by factorising in $\mathbf{Z}[\zeta_3]$ and $\mathbf{Z}[\sqrt{-2}]$ respectively, which are unique factorisation domains. The equation $x^3 + 1 = y^2$ is more difficult, and requires an infinite descent. The full argument is in [Cox89], and the conclusion is that all the solutions are given by the following table:

(X, Y)	(a, b)	$\gamma_2(\tau_0)$
$(0, 0)$	$(0, 0)$	0
$(-1, 0)$	$(2, 4)$	$-2^5 \cdot 3$
$(1, 2)$	$(-2, 8)$	$-2^5 \cdot 3 \cdot 5 \cdot 11$
$(1, -2)$	$(-2, 0)$	-2^5
$(2, 6)$	$(-4, 28)$	$-2^6 \cdot 3 \cdot 5 \cdot 23 \cdot 29$
$(2, -6)$	$(-4, 4)$	$-2^6 \cdot 3 \cdot 5$

To conclude, we note that all values of $\gamma_2(\tau_0)$ are accounted for by the values we computed for the imaginary quadratic fields we know. Notice that this value uniquely determines the imaginary quadratic field K, as they determine the j-invariant of an elliptic curve E, which in turn determines K as its ring of endomorphisms $\mathrm{End}_{\mathbf{Q}}(E)$. □

15 Analogues for other number fields

Now that we have a good understanding, both theoretical and practical, of abelian extensions of imaginary quadratic fields, we are naturally led to wonder what can be generalised to other number fields. Shimura [Shi70] develops a theory of complex multiplication for abelian varieties, which yields results as above for so called *CM fields*, which are quadratic imaginary extensions of totally real number fields. However, in general such a theory is currently unknown.

15.1 Stark's conjectures.

The simplest open case is that of quadratic real fields. Let $K = \mathbf{Q}(\sqrt{\Delta})$ with $\Delta > 0$ a fundamental discriminant, then there are a number of conjectural approaches to explicit class field theory for K. Perhaps the most celebrated is that provided by *Stark's conjectures*, of which we will now briefly discuss a particular case. In practice, it leads to a construction of the Hilbert class field of any real quadratic field following Cohen–Roblot [CR99].

Let us assume that the class number of K is greater than 1. Let H/K be the Hilbert class field of K, and choose a real embedding $\sigma : K \hookrightarrow \mathbf{R}$. Suppose that we have a quadratic extension F/H which is unramified at σ, but ramified at $\overline{\sigma}$, which is an abelian extension of K of conductor $\mathfrak{m}$. Call $G = \mathrm{Gal}(F/K)$ its Galois group, and let

$$\varphi_F : I_{\mathfrak{m}} \longrightarrow G$$

be the Artin map attached to F. The idea is now to choose an element of G, and define partial Dedekind zeta functions for the field K which capture some of the arithmetic of F. More precisely, choose $g \in G$ and define

$$\zeta_{K,g}(s) \sum_{0 \neq \mathfrak{a} \trianglelefteq \mathcal{O}_K, \varphi(\mathfrak{a}) = g} \mathrm{Nm}_{K/\mathbf{Q}}(\mathfrak{a})^{-s}.$$

This Dirichlet series converges absolutely for all $s \in \mathbf{C}$ with $\mathrm{Re}(s) > 1$ to a holomorphic function. It has meromorphic continuation to all of $\mathbf{C}$, with a simple pole at $s = 1$, and a simple zero at $s = 0$ (because the rank of $\mathcal{O}_K^\times$ is 1). Stark made a series of striking conjectures, which in this simple case would predict the following:

Conjecture 15.1. *There exists a unit $\varepsilon \in F$ such that $\sigma(\varepsilon) = e^{-2\zeta'_{K,g}(0)}$. Furthermore, $H = K(\varepsilon + \varepsilon^{-1})$.*

This conjecture suggests that it should be enough to gather enough information about such an extension F so as to compute $\zeta'_{K,g}(0)$ sufficiently accurately. This would allow one to construct the unit ε explicitly, and hence find the Hilbert class field H/K explicitly as a consequence.

First, note that we can easily find such a F by choosing $\alpha \in K$ such that $\sigma(\alpha) > 0$ but $\overline{\sigma}(\alpha) < 0$, and setting $F = H(\sqrt{\alpha})$. As the existence of F is guaranteed, we now turn to computing the value

$\zeta'_{K,g}(0)$. We will need Dirichlet L-functions attached to F/K, defined by

$$L_F(\chi, s) = \prod_{\mathfrak{p} \in I_\mathfrak{m}} (1 - \chi(\mathfrak{p})\mathrm{Nm}(\mathfrak{p})^{-s})^{-1},$$

where $\chi : G \to \mathbf{C}$ is a character of G. In other words, if we view χ as a character of $I_\mathfrak{m}$ via the natural quotient $I_\mathfrak{m} \to G$, and extend its definition to the set of all ideal by setting $\chi(\mathfrak{p}) = 0$ when $\mathfrak{p}$ ramifies in F/K, we see that this is just the usual Dirichlet L-series attached to χ. Call $\widehat{G}$ the group of characters $G \to \mathbf{C}$, then we have by orthogonality that

$$\zeta_{K,g}(s) = \deg(F/K)^{-1} \sum_{\chi \in \widehat{G}} L_F(\chi, s)\overline{\chi}(g).$$

This rewrites the ζ-function whose special value we are after as a sum of Dirichlet L-series we can get our hands on. Indeed, Dirichlet L-functions have a functional equation which imply

$$L'_F(\chi, 0) = 0, \qquad \text{when } \chi(\tau) = 1,$$

where $\mathrm{Gal}(F/H) = \langle \tau \rangle$. On the other hand, when $\chi(\tau) = -1$ it is not hard to prove that the conductor of χ is equal to $\mathfrak{m}$. This means that if we can efficiently compute the value of the derivatives of all Dirichlet L-functions of conductor $\mathfrak{m}$, then we would be able to construct the conjectural Stark units that generate F, and hence find H. Note that in practice, once such a unit is found, we may relatively straightforwardly verify that this is indeed the Hilbert class field, at which point we can forget all about the conjectural nature of Stark units that led us to it! This is precisely what Cohen–Roblot do, and their paper contains many beautiful ideas for computing special values of Dirichlet L-series. Rather than explaining how they work, we refer to their paper, and copy some examples found in *loc. cit.* :

16 Exercises

1. Let E be an elliptic curve such that $\mathrm{End}_{\mathbf{Q}}(E) \simeq K$ is an imaginary quadratic field. Show that an elliptic curve E' is isogenous to E if and only if $\mathrm{End}_{\mathbf{Q}}(E') \simeq K$.

2. Let $\mathcal{O}$ be an order in an imaginary quadratic field K, and $\mathfrak{a}$ a proper $\mathcal{O}$-ideal. Show that $j(\mathfrak{a})$ is real if and only if $\mathfrak{a}^2$ is principal.

3. Let $\mathcal{O}$ be the order of discriminant Δ in an imaginary quadratic field K, and $\mathfrak{a}$ a proper $\mathcal{O}$-ideal. Show that $\mathbf{Q}(j(\mathfrak{a}))$ is Galois over $\mathbf{Q}$ if and only if there is precisely one form in any genus of quadratic forms of discriminant Δ.

4. Let F/K be the abelian extension generated by $j(z)$ for all $z \in K$ in the upper half plane. Show that it corresponds to the subgroup
$$\mathbf{A}_{\mathbf{Q}}^{\times} K^{\times} \mathbf{A}_K^{f,\times}.$$

5. Find the Hilbert class field of $\mathbf{Q}(\sqrt{-23})$.

6. Use the theory of complex multiplication to make explicit the criterion for when a prime is of the form $x^2 + ny^2$, developed in chapter I, for $n = 11, 20$, and 101 (please use a computer).

7. Let E be an elliptic curve with CM by K. Prove that for any integral ideal $\mathfrak{m}$ in K, there exists a point t on E such that
$$[\alpha]t = 0, \text{ for } \alpha \in \mathcal{O}_K \iff \alpha \in \mathfrak{m}.$$
Use the main theorem of complex multiplication to show that $K_{\mathfrak{m}} = K(j(E), h_E(t))$. This generalises a result from the main text to non-principal ideals.

8. Consider $E : y^2 = x^3 + 1$, and let $K = \mathbf{Q}(\sqrt{-3})$. For each $N \geq 1$, let $K_N = K(h_E(E[N]))$ and $L_N = K(E[N])$. Calculate all of these fields for $N \leq 4$, and show that they are abelian extensions of K.

9. Show that the denominators appearing in the power series X, Y defining the Tate uniformisation are essential, by explaining why the projective coordinates
$$\big((1-u)^3 X(u,q) \ : \ (1-u)^3 Y(u,q) \ : \ (1-u)^3\big)$$
do **not** define a morphism of schemes over $R = \mathbf{Z}[\![q]\!]$ from $\mathbf{G}_m/R$ to $\mathrm{Tate}(q)$.

10. Let K be a quadratic number field with ring of integers $\mathcal{O}_K$ of discriminant d_K. Prove that if $\mathcal{O}$ is the order of conductor f, then we have
$$h(\mathcal{O}) = \frac{h(\mathcal{O}_K) f}{[\mathcal{O}_K^{\times} : \mathcal{O}^{\times}]} \cdot \prod_{p \mid f} \left(1 - \left(\frac{d_K}{p}\right) \frac{1}{p}\right).$$

11. Show that there are precisely 13 discriminants $\Delta < 0$ such that $h(\Delta) = 1$.

Chapter III

Modular Curves

In this chapter, we will see how Heegner's solution to the class number one problem has a more conceptual geometric interpretation, in terms of rational points on non-split Cartan modular curves $X_{\mathrm{ns}}^{+}(N)$. The proof of Heegner is essentially an explicit determination of all the rational points on the genus 1 curve $X_{\mathrm{ns}}^{+}(24)$. This leads us very naturally to one of the oldest questions in mathematics:

Question: Given an algebraic curve $\mathcal{C}$ over $\mathbf{Q}$, can we describe $\mathcal{C}(\mathbf{Q})$?

As you know, this is an extremely difficult question, which remains today one of the central questions in arithmetic geometry. Some successful methods have been established in the previous century, and continue to be developed today. We will discuss descent via isogenies for abelian varieties, as well as the p-adic method of Chabauty–Coleman and some of its modern generalisations. Finally, we return to our initial setting and apply these techniques to the study of rational points on modular curves, where the moduli interpretation both yields a lot of extra information, as well as concrete interpretations of the results on rational points. As an application, we discuss the proof of Mazur–Tate of the fact that no elliptic curve over $\mathbf{Q}$ has a rational point of order 13.

17 Class number one and $X_{\mathrm{ns}}^{+}(N)$

We ended the previous chapter with a discussion of Heegner's proof of the class number one problem, as interpreted by Stark and Birch. The proof consisted of an application of the main theorem of CM theory, leading to the Diophantine equation

$$2X(X^3+1) = Y^2,$$

obtained via a series of algebraic manipulations using the relations between Weber functions developed above. Serre gave a more conceptual geometric interpretation of this proof, which we will now present.

Define a *non-split Cartan subgroup* of $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$ to be the image of $(\mathcal{O}/N\mathcal{O})^\times$ acting on $(\mathcal{O}/N\mathcal{O})$ via multiplication, where $\mathcal{O}$ is an order in an imaginary quadratic field such that every prime p dividing N is inert in $\mathcal{O}$. All non-split Cartan subgroups are conjugate to each other, so we choose one and call it $C_{\mathrm{ns}}(N)$. The normaliser $C_{\mathrm{ns}}^+(N)$ of $C_{\mathrm{ns}}(N)$ is generated by a ν-tuple of involutions, where ν is the number of prime divisors of N, and it has size

$$|C_{\mathrm{ns}}^+(N)| = 2^\nu N^2 \prod_{p|N} \left(1 - \frac{1}{p^2}\right).$$

We will use this subgroup to define a modular curve $X_{\mathrm{ns}}^+(N)$ over $\mathbf{Q}$, as follows. We start with the modular curve $X(N)$ with full level $\Gamma(N)$-structure, parametrising elliptic curves, together with a basis for its N-torsion. This curve is *disconnected*, as the Weil pairing attaches to any such basis a primitive N-th root of unity, and in fact $X(N)$ is the disjoint union, indexed by $(\mathbf{Z}/N\mathbf{Z})^\times$, of curves which are isomorphic over $\mathbf{C}$ to $\Gamma(N)\backslash\mathfrak{H}^*$. Now, for any subgroup H of $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$ we may define the algebraic curve

$$X_H = H\backslash X(N).$$

Whenever the determinant map $\det : H \to (\mathbf{Z}/N\mathbf{Z})^\times$ is surjective, the curve X_H is a connected algebraic curves, defined over $\mathbf{Q}$. The subgroup $H = C_{\mathrm{ns}}^+(N)$ has this property, so we obtain an algebraic curve $X_{\mathrm{ns}}^+(N)$ over $\mathbf{Q}$ with the following properties:

- Moduli interpretation: It parametrises pairs (E, φ), where E is an elliptic curve, and $\varphi : E[N] \overset{\sim}{\to} (\mathbf{Z}/N\mathbf{Z})^2$ is an isomorphism, taken up to the equivalence relation: $(E,\varphi) \sim (E', \varphi')$ whenever $E = E'$ and $\varphi^{-1} \circ \varphi' \in C_{\mathrm{ns}}^+$,
- The forgetful projection map $X_{\mathrm{ns}}^+(N) \to X(1)$ is of degree $N\varphi(N)/2^\nu$.

We now come to Serre's geometric approach to the class number one problem:

Theorem 17.1. *Let K be an imaginary quadratic field with class number one, and N an integer such that any prime divisor of N is inert in K. Then any elliptic curve with CM by $\mathcal{O}_K$ gives rise to a unique rational point on $X_{ns}^+(N)$.*

Proof. Consider the elliptic curve E over $\mathbf{Q}$ with CM by $\mathcal{O}_K$, unique up to isomorphism over $\mathbf{C}$. The natural actions of the Galois group of $\mathbf{Q}$ and the endomorphism ring of E on $E[N]$ give us morphisms

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z}), \quad (\mathcal{O}_K/N\mathcal{O}_K)^\times \to \mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z}).$$

By definition, the image of $(\mathcal{O}_K/N\mathcal{O}_K)^\times$ is a non-split Cartan subgroup $C_{\mathrm{ns}}(N)$. The Galois action of the subgroup $\mathrm{Gal}(\overline{Q}/K)$ commutes with the endomorphism ring $\mathcal{O}_K$, so that the image of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ must lie in the normaliser $C_{\mathrm{ns}}^+(N)$. This shows that K gives rise to a unique rational point on $X_{\mathrm{ns}}^+(N)$. □

This means that we obtain a beautiful geometric approach to the class number one problem! Indeed, if we manage to find the rational points on $X_{\text{ns}}^{+}(N)$ for some N, then we have found all imaginary quadratic fields K with class number one such that all the prime divisors of N are inert in K. This condition is not really an obstacle, and a determination of the rational points on $X_{\text{ns}}^{+}(N)$ for any N will provide us with a complete solution to the class number one problem. The reason has already appeared in Chapter I: If K has class number one, then any prime less than $(1 - \Delta_K)/4$ is inert in K.

The only question remaining is how Heegner's proof fits into this strategy. We will leave this to the reader's imagination, saying only that the non-split Cartan modular curve $X_{\text{ns}}^{+}(24)$ has genus 1, just like the projective curve defined by the Diophantine equation considered by Heegner. Furthermore, we know that the modular function $\mathfrak{f}^2$ is invariant under $\Gamma(24)$, and in fact it is even invariant under an index 3 subgroup of $\Gamma_{\text{ns}}^{+}(24)$. This might convince you that Heegner's arguments take place on modular curves closely related to $X_{\text{ns}}^{+}(24)$, but perhaps Serre's reinterpretation is slightly cleaner.

Remark. It has been shown that all rational points on $X_{\text{ns}}^{+}(N)$ come from CM elliptic curves, whenever this curve has genus at least 1, **except** when $N = 13$. The genus 3 curve $X_{\text{ns}}^{+}(13)$ has proven stubbornly resistant to current techniques, and is affectionately known as the *cursed modular curve*. It seems likely that the recent method of Chabauty–Coleman–Kim provides a natural strategy for finding all the rational points on this last remaining open case. Time will tell.

17.1 The image of Galois representation $E[p]$.

After giving the argument that the image of the Galois representation on $E[N]$ for an elliptic curve with CM by a field in which all primes dividing N are inert in p, we should mention the general classification of maximal subgroups in $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$, and Serre's big image theorem. [+++]

18 Rational points on curves

Let $X_{\mathbf{Q}}$ be a smooth projective curve, then the search for $X(\mathbf{Q})$ is one of the oldest problems in mathematics. The arithmetic is heavily influenced by topological information. When $g \leq 1$, have good methods for finding rational points, given one initial point.

Abel-Jacobi map.

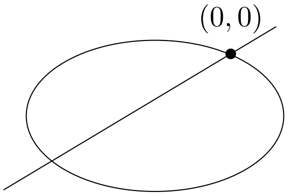
18.1 Curves of genus 0.

The first fundamental question to ask is: Are there any rational points at all? For projective curves of genus 0, there is an easy way to decide whether this is true, via *Hilbert symbols*. This is a consequence of the fact that the Hasse principle holds for curves of genus 0, which states that such a curve has a point over $\mathbf{Q}$ if and only if it has a point over $\mathbf{R}$, and over every p-adic field $\mathbf{Q}_p$. [+++]

Once we have found a rational point, there must be infinitely many such points, parametrised by the set of lines with rational slope going through this point, by considering its second intersection point with the conic. This is perhaps best explained on an example. Suppose we would like to find the integral solutions to the Diophantine equation

$$3x^2 + 4y^2 = 7z^2, \qquad x, y, z \in \mathbf{Z}$$

This is of course equivalent to finding rational solutions on the conic $3X^2 + 4Y^2 = 7$ defined over $\mathbf{Q}$, which has an obvious point $(1, 1)$. Changing coordinates so that this point becomes the origin, the equation becomes $3X_0^2 + 6X_0 + 4Y_0^2 + 8Y_0 = 0$, which are parametrised by the lines with rational slope through $(0, 0)$.



A straightforward calculation now reveals that the full set of integral solutions to the original equation is

$$\{\lambda \cdot (1,1,1)\} \cup \left\{\mu \cdot (4\nu^2 - 8\nu - 3, -4\nu^2 - 6\nu + 3, 4\nu^2 + 3) \ : \ \mu, \nu \in \mathbf{Z}\right\}.$$

We see that the theory of rational points on genus 0 curves is extremely well understood, even on an algorithmic level. Of course, certain issues remain, such as finding a rational point once its existence has been proven by the Hasse principle, but we won't address such issues here.

18.2 Curves of genus 1.

For curves of genus 1, the theory becomes much more mysterious. Already the decision whether or not a projective genus 1 curve has a point is very difficult, as the Hasse principle fails in this setting. Historically, the first counterexample was given by Selberg, who showed that

$$3x^3 + 4y^3 + 5z^3 = 0,$$

has solutions in $\mathbf{R}$ and in every p-adic field $\mathbf{Q}_p$ (we say it has solutions *everywhere locally*) but it has no rational points over $\mathbf{Q}$! See also the exercises.

Suppose that our curve of genus 1 is known to have a rational point, then it is an elliptic curve. The Mordell–Weil theorem, whose proof is discussed in the next section, guarantees that

$$E(\mathbf{Q}) \simeq E_{\text{tors}}(\mathbf{Q}) \times \mathbf{Z}^r,$$

for some integer r called the *rank* of E over $\mathbf{Q}$, and some finite group $E_{\text{tors}}(\mathbf{Q})$. The Mordell–Weil theorem in fact proves this also when $\mathbf{Q}$ is replaced by any number field K, and E is replaced by any abelian variety defined over K. For the special case of elliptic curves over $\mathbf{Q}$, much more specific information is available. Whereas the rank remains a difficult and mysterious invariant, the torsion is a finite abelian group whose order is extremely constrained. More precisely, it was shown by Mazur in [+++] what all the various possibilities are.

Theorem 18.1 (Mazur). *TODO.*

This theorem is extremely difficult to prove, and already special cases of it require a lot of hard work. At the end of this course, we will show that C_{13} can never be a subgroup of $E_{\mathrm{tors}}(\mathbf{Q})$, requireing already some rather sophisticated machinery.

Of course the question remains how to compute $E(\mathbf{Q})$ explicitly, or say anything about it. The torsion is usually very easily computed, but the computation of the rank requires us to talk about the theory of descent, and will be very natural once we have discussed the proof of the Mordell–Weil theorem. It is striking that the theory of Selmer groups is both the only available to tool to show that the Mordell–Weil rank is finite, and the only available tool to compute any information about the rational points at all.

18.3 Curves of genus ≥ 2.

For curves of higher genus, the situation is radically different. For a long time, the only precise statement was a conjecture due to Mordell, which was finally proved by Faltings with a phenomenal argument that earned him the Fields medal:

Theorem 18.2 (Faltings 1983). *Let $X_{\mathbf{Q}}$ be of genus $g \geq 2$, then $X(\mathbf{Q})$ is finite.*

The proof of Faltings uses the theory of heights, and is very far from being constructive. In particular, it gives us no method to find all rational points on a given projective curve of genus ≥ 2, or even bound the number of such points in an effective way. In that sense, even after the work of Faltings the problem of determining rational points is in some sense still unsolved and poorly understood. An earlier approach due to Chabauty, which allowed him to prove Mordell's conjecture under an additional hypothesis, has the advantage that it can be made into an explicit method for finding bounds for the number of points, or even finding all the rational points explicitly in some favourable cases. This was noted by Coleman when he developed his p-adic theory of integration, and we will discuss this method in detail below. It should also be noted that generalisations of Chabauty's idea remain today a very active and exciting area of mathematics, which exploits many deep ideas from algebraic topology to address arithmetic questions, as done for instance in the work of Kim.

The section conjecture. The fact that the behaviour of rational points is fundamentally different for *hyperbolic curves* such as projective curves of genus ≥ 2 was realised and made precise by Grothendieck in his *section conjecture.* The theory of algebraic fundamental groups shows that there is a fundamental sequence

$$1 \longrightarrow \pi_1^{\text{ét}}(X_{\overline{\mathbf{Q}}}; b) \longrightarrow \pi_1^{\text{ét}}(X_{\mathbf{Q}}; b) \longrightarrow \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow 1.$$

Any rational point gives rise to a splitting of this short exact sequence, a fact which follows from the functoriality of algebraic fundamental groups. Grothendieck conjectures that if X is a hyperbolic curve, then in fact any splitting of this short exact sequence should come from a rational point. This is an extremely deep conjecture with far-reaching consequences, which seems to remain out of reach today.

19 Descent via isogeny for abelian varieties

In what follows, we will assume we are given an abelian variety A over a number field K.

19.1 Descent via isogeny.

Galois cohomology, Selmer groups, computations of 2-descent on elliptic curves, example of other isogeny (Dokchitser).

Let $\psi : A \to A'$ be an isogeny of degree p, and let $\psi^\vee : A' \to A$ be its dual isogeny. There is an exact sequence of G_K-modules

$$0 \to A'[\psi^\vee](\overline{\mathbf{Q}}) \to A'(\overline{\mathbf{Q}}) \xrightarrow{\psi^\vee} A(\overline{\mathbf{Q}}) \to 0 \tag{III.1}$$

Taking invariants, we extract an injection of $A(\mathbf{Q})/\psi^\vee(A'(\mathbf{Q}))$ into the cohomology group $\mathrm{H}^1(\mathbf{Q}, A'[\psi^\vee])$. If we want to prove finiteness of the former, it would be great if the latter were finite (and computable). However, this group is absolutely enormous, and actually the image lands in a very nice, finite, computable subgroup. This subgroup is the *Selmer group*, which is defined as follows. Looking at the various restriction maps between long exact sequences in Galois cohomology coming from (III.1), we obtain the following commutative diagram with exact rows:

$$\begin{tikzcd}
A(\mathbf{Q}) \arrow[r] \arrow[d] & \mathrm{H}^1(\mathbf{Q}, A'[\psi^\vee]) \arrow[r] \arrow[d] \arrow[dr, dashed, "\rho"] & \mathrm{H}^1(\mathbf{Q}, A) \arrow[d] \\
\prod_v A(\mathbf{Q}_v) \arrow[r] & \prod_v \mathrm{H}^1(\mathbf{Q}_v, A'[\psi^\vee]) \arrow[r] & \prod_v \mathrm{H}^1(\mathbf{Q}_v, A)
\end{tikzcd} \tag{III.2}$$

We define the *Selmer group* $\mathrm{H}^1_f(\mathbf{Q}, A'[\psi^\vee])$ to be the kernel of the map ρ. The *Shafarevich–Tate group* $Ш(\mathbf{Q}, A')$ is the kernel of the rightmost map of this diagram. We obtain from this definition the fundamental exact sequene

$$0 \to A(\mathbf{Q})/\psi^\vee(A'(\mathbf{Q})) \to \mathrm{H}^1_f(\mathbf{Q}, A[\psi^\vee]) \to Ш(\mathbf{Q}, A')[\psi^\vee] \to 0.$$

This sequence is of primordial importance. The main reason is that the Selmer group in the middle is finite, and somewhat computable.

Theorem 19.1. *The Selmer group* $\mathrm{H}^1_f(\mathbf{Q}, A'[\psi^\vee])$ *is finite.*

Proof. We will sketch the argument, see [+++] for a full proof. Let S be the set of places consisting of

- the archimedean places,
- the places of bad reduction of A,
- the places dividing the degree of ψ.

Now define the set $\mathrm{H}^1(K, A'[\psi^\vee]; S)$ to be the classes in $HH^1(K, A'[\psi^\vee])$ that are unramified outside S. The set $\mathrm{H}^1(K, A'[\psi^\vee]; S)$ is finite, because of the Dirichlet unit theorem and finiteness of class groups. □

A very important consequence is that the quotient on the left is a finite group! This is known as the *weak Mordell–Weil theorem*, and it is the first step in proving that $A(\mathbf{Q})$ is a finitely generated abelian group. The theory of heights may be used to show that the finiteness of the quotient implies the theorem, but we will not discuss this here. For more details, see [+++]. We note also that it follows from the above theorem that $\text{Ш}(\mathbf{Q}, A')[\psi^\vee]$ is finite. It is a famous open conjecture that in fact all of $\text{Ш}(\mathbf{Q}, A')$ is finite.

19.2 The Mordell–Weil theorem.

The theory of heights has long been a crucial tool in arithmetic geometry, but any serious discussion of it would lead us too far. To motivate it, choose your favourite elliptic curve over $\mathbf{Q}$ and pick a rational point P on it. Now use a computer algebra package to print the x-coordinate of mP for growing m. The digits will look like they trace out the area under a parabola. This is what the theory of heights explains!

More precisely, let A_K be an abelian variety over a number field K as before, then the theory of heights provides us for every choice of a very ample symmetric line bundle on A with a function

$$h : A(K) \longrightarrow \mathbf{R},$$

such that for any constant C, the set $\{P \in A(K) \ : \ h(P) \leq C\}$ is finite. Moreover:

- Let $Q \in A(K)$, then there is a constant C_Q so that $h(P+Q) \leq 2h(P)+C_Q$, for all $P \in A(K)$.
- There is an integer $m \geq 2$ and a constant C_A, so that $h(mP) \geq m^2 h(P) - C_A$ for all $P \in A$.

Just the mere existence of such functions implies the Mordell–Weil theorem.

Theorem 19.2 (Mordell–Weil). *Let K be a number field and A_K an abelian variety. Then the group $A(K)$ is finitely generated.*

Proof. let $m \geq 2$ be any integer, and let $h : A(K) \to \mathbf{R}$ be a height function as above. Considering the multiplication by m isogeny $[m] : A \longrightarrow A$, we know that

$$A(K)/mA(K) \hookrightarrow \mathrm{H}^1_f(K, A[m]),$$

and so $A(K)/mA(K)$ is finite by the finiteness of Selmer groups. Choose lifts $Q_1, Q_2, \ldots, Q_r \in A(K)$ of all the elements in $A(K)/mA(K)$, and choose an arbitrary $P_0 \in A(K)$. Recursively choose P_n for $n \geq 1$ satisfying the relation

$$P_{n-1} = mP_n + Q_{i_n}, \qquad \text{for some } \; i \leq i_n \leq r.$$

Now for any n, we have

$$h(P_n) \leq \frac{1}{m^2}(h(mP_n) + C_A) \leq \frac{1}{m^2}(2h(mP_{n-1}) + C_Q + C_A) \leq \ldots \leq 2^{-n}h(P) + (C_Q + C_A)/2,$$

where C_Q is the maximum of the constants C_{Q_i}. The constants C_Q and C_A are independent of P. When n becomes very large, the height of P_n is bounded by, say, $1 + (C_Q + C_A)/2$, and there are only finitely many such points. This means that P is a linear combination of these finitely many points and the finitely many Q_i, and so $A(K)$ is finitely generated. □

20 Computing Mordell–Weil groups of elliptic curves

20.1 Computing $E(\mathbf{Q})_{\text{tors}}$.

Hasse-Weil

20.2 Computing Selmer groups.

Selmer group of isogeny is explicit (Stoll Schaeffer) Kummer theory for full 2-descent. Local methods in appendix?

21 The method of Chabauty–Coleman

Example. This example is taken from McCalum–Poonen [+++] and arose when studying [+++]. Consider the genus 2 hyperelliptic curve of rank 1

$$X : y^2 = x^6 + 8x^5 + 22x^4 + 22x^3 + 5x^2 + 6x + 1,$$

First, note that $X(\mathbf{Q}) \supseteq \{\infty^{\pm}, (-3, \pm 1), (0, \pm 1)\}$. We will try to prove that these are in fact all the rational points, using the method of Chabauty–Coleman. Choose $p = 3$, and observe that $X(\mathbf{F}_3) = \{\infty^{\pm}, (0, \pm 1)\}$. We can use the residue disk of $(0, 1)$ to determine an annihilating differential ω_0 with just one digit of precision:

$$\omega_0 \equiv x\frac{dx}{y} \equiv (x - x^3 + \ldots)dx \pmod 3.$$

Now run through all residue disks, for instance $(0,-1)$, where

$$\int_{(0,-1)}^{z} \omega_0 = x^2 \cdot (a_0 + a_1 x + a_2 x + \ldots)\,,\ v_3(a_i) \geq 0,\ \text{for}\ \ i \leq 2$$

has at most 2 zeroes $z \in 3\mathbf{Z}_3$ by Newton polygon estimates. This shows that

$$X(\mathbf{Q}) = \left\{\infty^{\pm}, (-3, \pm 1), (0, \pm 1)\right\}.$$

21.1 The method of Chabauty–Coleman–Kim.

In practice, the p-adic method of Chabauty–Coleman is often applicable, and has enabled us to find all rational points on large numbers of explicit examples. However, there are limitations to this method, most prominently the condition $r < g$. Recently, Minhyong Kim has proposed a deep and far-reaching non-abelian generalisation of the method of Chabauty–Coleman. Its technical backbone is formidable, and its successful execution on explicit examples largely remains to be done, though a good number of recent instances have appeared in the literature. We will try to give the reader a taste of some of the main characters in the theory, remaining largely vague.

In the case of elliptic curves, we classified torsors of the p-adic Tate module $T_p(E)$, or a finite quotient $E[p]$ of it.

- The resulting classifying space $\mathrm{H}^1(G_{\mathbf{Q}}, E[2])$ was a Galois cohomology group, amenable to explicit computation.
- The torsors attached to rational points satisfied a set of *Selmer* finiteness conditions.

For a hyperbolic curve, we could try the analogue:

$$\kappa_b : X(\mathbf{Q}) \longrightarrow \mathrm{H}^1(G_{\mathbf{Q}}, \pi_1^{\text{ét}}\left(X_{\overline{\mathbf{Q}}}, b)\right),$$

which according to the section conjecture should even be an isomorphism! Compare this to the analogous statement for elliptic curves:

$$\kappa_b : \widehat{E(\mathbf{Q})} \xrightarrow{\sim} \mathrm{H}^1_f\left(G_{\mathbf{Q}}, \pi_1^{\text{ét}}(E_{\overline{\mathbf{Q}}}, b)\right).$$

For a hyperbolic curve, the classifying set of $\pi_1^{\text{ét}}$ is too wild to have any extra structure. We restrict to the $\mathbf{Q}_p$-pro-unipotent fundamental group, by considering the category of locally constant $\mathbf{Q}_p$-sheaves on $X_{\overline{\mathbf{Q}}}$ which are successive extensions of the constant sheaf $\mathbf{Q}_p$ (unipotent). Tensor compatible fibre functor isomorphisms yield $\mathrm{U}^{\text{ét}} := \pi_1^{u,\text{ét}}(X_{\overline{\mathbf{Q}}}; b)$ and étale path torsors $\pi_1^{u,\text{ét}}(X_{\overline{\mathbf{Q}}}; b, z)$, as well as finite level versions $\mathrm{U}_n^{\text{ét}}$ coming from n-step unipotent sheaves. These torsors come with an action of Galois groups G_T. This gives us a series of associations $X(\mathbf{Q}) \to \mathrm{H}^1(G_{\mathbf{Q}}, \mathrm{U}_n)$, and the image lies in a subspace $\mathrm{H}^1_f(G_{\mathbf{Q}}, \mathrm{U}_n)$ defined by certain finiteness conditions:

1. Unramified outside $S \cup \{p\}$ where S is the set of primes of bad reduction of X.

2. Crystalline at p, defined using non-abelian p-adic Hodge theory.

Amazingly, $\mathrm{H}^1_f(G_{\mathbf{Q}}, \mathrm{U}_n)$ has the structure of an algebraic variety over $\mathbf{Q}_p$! We call it the *Selmer variety* of level n. All of this is also true over $\mathbf{Q}_p$ by restriction, get local versions.

This is all still very abstract! Like in the case of Chabauty–Coleman, we will need a clear connection with a much more explicit de Rham realisation.

1. Consider the category of (n-step) unipotent vector bundles with integrable connection, fibre functor yields $\mathrm{U}_n^{\mathrm{dR}} := \pi_1^{\mathrm{dR}}(X_{\mathbf{Q}_p}, b)_n$ and de Rham path torsors $\pi_1^{\mathrm{dR}}(X_{\mathbf{Q}_p}; b, z)_n$. Come with a Hodge filtration and Frobenius structure.

2. There is a non-abelian Dieudonné functor $\mathbf{D}_n : \mathrm{H}^1_f(G_{\mathbf{Q}_p}, \mathrm{U}_n^{\text{ét}}) \to \mathrm{U}_n^{\mathrm{dR}}/\mathrm{Fil}^0\mathrm{U}_n^{\mathrm{dR}}$, given on torsors by
$$\operatorname{Spec} \mathcal{P} \mapsto \operatorname{Spec}(\mathcal{P} \otimes \mathrm{B}_{\mathrm{cris}})^{G_{\mathbf{Q}_p}}.$$

Now get the diagram:

$$\begin{tikzcd}
X(\mathbf{Q}) \arrow[r] \arrow[d, "j_n^{\text{ét}}"'] & X(\mathbf{Q}_p) \arrow[d, "j_{n,p}^{\text{ét}}"'] \arrow[rd, "j_n^{\mathrm{dR}}"] & \\
\mathrm{H}^1_f(G_T, \mathrm{U}_n^{\text{ét}}) \arrow[r, "\mathrm{loc}_{n,p}"] & \mathrm{H}^1_f(G_p, \mathrm{U}_n^{\text{ét}}) \arrow[r, "\mathbf{D}_n"'] & \mathrm{U}_n^{\mathrm{dR}}/\mathrm{Fil}^0\mathrm{U}_n^{\mathrm{dR}}
\end{tikzcd}$$

This diagram commutes, and moreover j_n^{dR} is given by iterated Coleman integrals! It is highly transcendental in nature, but has a much more computable flavour.

Some fundamental questions remain:

1. Is $\dim \mathrm{H}^1_f(G_T, \mathrm{U}_n^{\text{ét}}) < \dim \mathrm{U}_n^{\mathrm{dR}}/\mathrm{Fil}^0\mathrm{U}_n^{\mathrm{dR}}$?

2. Can we describe image of Selmer varieties explicitly?

It can be shown that an affirmative answer to the first question would be implied by the Bloch–Kato or Fontaine–Mazur conjecture. Even assuming this, it is not clear how to answer the second question, though many special cases have been worked out by Balakrishnan, Besser, Dan-Cohen, Dogra, Kedlaya, and Wewers.

22 Rational points on modular curves

We now have two general methods for finding rational points on curves: Descent via isogeny on the Jacobian, and the Chabauty–Coleman method. As our interest in finding rational points started in a situation where the curve in question was a modular curve, we wonder if more specific techniques apply. After all, modular curves are extremely special, and apart from their many miraculous geometric properties, they also come with a moduli interpretation, which facilitates our task in many cases. There are two distinguished classes of rational points: Cusps and CM points.

22.1 Calculations with cusps.

Recall that for any subgroup $H \leq \mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$ we have defined a moduli problem $\mathcal{P}_H$ which is relatively representable and and étale over $\mathrm{Ell}/\mathbf{Z}[1/N]$. The corresponding scheme $Y_{\mathcal{P}_H}$ is defined over $\mathbf{Q}(\mu_N)^{\det H}$. We may wonder what the rationality properties of the cusps are. The following theorem may be found in Katz–Mazur [+++]

Theorem 22.1. *Cusps of $Y_{\mathcal{P}_H}$ correspond to $\mathcal{P}_H$-level structures on $\mathrm{Tate}(q)$ over $\mathbf{Z}\llbracket q^{1/N}, \zeta_N\rrbracket$ up to the automorphisms $q^{1/N} \mapsto \zeta_N^a q^{\pm 1/N}$ and the automorphism given by multiplication by -1 on the curve. This correspondence is equivariant under the natural actions of $\mathrm{Gal}(\mathbf{Q}(\mu_N)/\mathbf{Q})$ on both sides.*

Of course, you were already taught how to count cusps of modular curves *group theoretically* by conjugating matrices and counting indices of subgroups. This all happened on the Riemann surface $\mathfrak{H}$ and gave you no understanding of the arithmetic properties of the cusps. Let us do some examples to convince you further that you have learned something new.

Example. Suppose $H = \Gamma_1(5)$, then the cusps correspond to points of order 5 on $\mathrm{Tate}(q)$ over $\mathbf{Z}\left[\frac{1}{5}, \zeta_5\right]\llbracket q^{1/5}\rrbracket$ which are the images of $q^{a/5}\zeta_5^b$ with $a, b \in (\mathbf{Z}/5\mathbf{Z})^2\backslash(0,0)$. Keeping track of automorphisms, we find that there are 4 cusps

$$\{\zeta_5, \zeta_5^4\}, \ \{\zeta_5^2, \zeta_5^3\}, \ \{q^{\pm 1/5}\zeta_5^a\}, \ \{q^{\pm 2/5}\zeta_5^a\}.$$

The Galois group $\mathrm{Gal}(\mathbf{Q}(\mu_5)/\mathbf{Q})$ acts through its C_2-quotient by swapping the first two cusps and fixing the last two. Therefore there are 2 cusps defined over $\mathbf{Q}(\mu_5)^+$ which are interchanged under Galois, and two cusps defined over $\mathbf{Q}$.

Example. Suppose $H = \Gamma_0(p)$ for some prime p. Then cusps correspond to subgroups of order p on $\mathrm{Tate}(q)$, of which there are precisely $p+1$ and they are of the form $\langle \zeta_p^a q^{b/p}$, modulo automorphisms $q^{1/p} \mapsto \zeta_p^a q^{\pm 1/p}$. This means there are two cusps corresponding to

$$\{\langle \zeta_p \rangle\} \quad \text{and} \quad \{\langle q^{1/p}\zeta_p^a \rangle\}_{0 \leq a \leq p-1}.$$

We also see that both cusps are defined over $\mathbf{Q}$.

Example. non-split Cartan (Rebolledo–Wuthrich).

22.2 Calculations with CM points.

Heegner points

Examples of equations, for $X(3), X_1(5)$ and $X_1(11)$. Try a 5-descent on $X_1(11)$. Say this is bad in general. Lay groundwork for solutions by pure thought: Symmetry, bad reduction, large torsion, etc.

23 Rational points on $X_1(13)$

Let $E_{\mathbf{Q}}$ be an elliptic curve, then the Mordell–Weil theorem states that

$$E(\mathbf{Q}) \simeq E(\mathbf{Q})_{\text{tors}} \times \mathbf{Z}^r,$$

where $E(\mathbf{Q})_{\text{tors}}$ is a finite abelian group we call the *torsion subgroup* of $E(\mathbf{Q})$, and r is a non-negative integer called the *rank* of $E(\mathbf{Q})$. Computing the torsion subgroup on any given concrete example is

23.1 A 19-descent on $J_1(13)$.

23.2 Grothendieck topologies and fppf cohomology.

23.3 Mazur's Theorem on torsion.

24 Exercises

1. Show that an elliptic curve $E_{\mathbf{Q}}$ has a rational 3-isogeny if and only if it has a model $y^2 = x^3 + a(x-b)^2$ for some $a, b \in \mathbf{Q}$. Formulate an analogous condition for E to have a rational 3-torsion point.

2. Suppose that $E_{\mathbf{Q}}$ has good reduction at 5. Show that $E(\mathbf{Q})[p] = 0$ for every prime $p \geq 11$.

3. Compute $E(\mathbf{Q})_{\text{tors}}$ for $E_1 : y^2 = x^3 + 4x$ and $E_2 : y^2 = x^3 + 4$.

4. Show that $E : y^2 = x^3 - 49x$ has Mordell–Weil rank 1.

5. Find all rational solutions to the equation $y^2 = x^3 - x$.

6. Prove that two isogenous elliptic curves have the same Mordell–Weil rank.

7. (Fermat) Prove that $E : y^2 = x^3 - 4x$ has $E(\mathbf{Q}) \simeq \mathrm{C}_2 \times \mathrm{C}_2$. Use this to show that $a^4 + b^4 = c^4$ has no non-trivial integer solutions.

8. Let $E : y^2 = x^3 + ax + b$ with $a, b \in \mathbf{Q}$. For a squarefree integer $d \neq 1$ set $E_d : y^2 = x^3 + ad^2x + bd^3$ the *quadratic twist* of E by d. Prove that E and E_d are isomorphic over $\mathbf{Q}(\sqrt{d})$, but not over $\mathbf{Q}$. Finally, show that

$$\text{rank } E(\mathbf{Q}) + \text{rank } E_d(\mathbf{Q}) = \text{rank } E(\mathbf{Q}(\sqrt{d})).$$

9. Prove that the Hasse principle fails in genus 1 by showing that Selberg's example

$$C : \ 3x^3 + 4y^3 + 5z^3 = 0$$

has points everywhere locally, while $C(\mathbf{Q}) = \emptyset$.

10. Show that $X_0(11)$ is an elliptic curve of rank 0. Find the number of rational points.
Hint: Do a descent via 5-isogeny on the forgetful map $X_1(11) \to X_0(11)$.

A. Galois cohomology

Restriction map, define unramified classes. Inflation restriction. Poitou-Tate.

Kummer Theory.

B. Elliptic curves

Definition. An elliptic curve is [+++]

We will only consider the case where S is the spectrum of a field k in this course, and as we will see the theory of elliptic curves is strongly influenced by the nature of k.

Let E_1 and E_2 be two elliptic curves over k, then we define a *homomorphism* to be a rational map $E_1 \to E_2$ defined over k that is also a homomorphism for the group law. Any rational map $E_1 \to E_2$ that sends the origin to the origin is automatically a homomorphism. We call an element

$$\lambda \in \mathrm{Hom}(E_1, E_2)$$

an *isogeny* if it satisfies one of the following three equivalent conditions:

- $\lambda \neq 0$,
- $\mathrm{Ker}(\lambda)$ is finite,
- λ is surjective.

If there exists an isogeny from E_1 to E_2, then there exists an isogeny from E_2 to E_1, and we say that E_1 and E_2 are *isogenous*, which defines an equivalence relation on the set of elliptic curves over k. Define $\mathrm{End}(E) = \mathrm{Hom}(E, E)$ to be the ring of endomorphisms of E, and set

$$\mathrm{End}_{\mathbf{Q}}(E) = \mathrm{End}(E) \otimes_{\mathbf{Z}} \mathbf{Q}.$$

For any field k, the ring $\mathrm{End}(E)$ is a free module of finite rank, and $\mathrm{End}_{\mathbf{Q}}(E)$ is a division ring of finite rank over $\mathbf{Q}$. Deuring determined all the possible isomorphism types of $\mathrm{End}_{\mathbf{Q}}(E_{\mathbf{C}})$, and showed that it is isomorphic to $\mathbf{Q}$, an imaginary quadratic field, or a quaternion algebra over $\mathbf{Q}$ ramified at a prime p and ∞. The last case can only occur when k has characteristic p.

Any elliptic curve

Definition, Tate modules, Mordell-Weil, endomorphism rings, isogenies, L-function.

Weil pairing. Shimura 4.3

Kummer pairing.

25 Elliptic curves over $\mathbf{C}$

Lattice, Weierstrass p.

26 Elliptic curves over $\mathbf{F}_p$

Frobenius, zeta function, Hasse-Weil, torsion injects.

27 Elliptic curves over $\mathbf{Q}_p$

Tate algorithm, formal groups.

C. Class Field Theory

We now recall some classical results on algebraic number theory, and review global class field theory from a practical point of view. We will not prove any of the statements, but instead concentrate on how to use the theory on concrete examples. You should aim to do as many exercises as possible. Our presentation of this material is modeled on notes of a Part III course taught by V. Dokchitser (unpublished) and an expository article on global class field theory by Poonen [Poo].

Let $K/\mathbf{Q}$ be number field, and $\mathcal{O}_K$ the integral closure of $\mathbf{Z}$ in K. The ring $\mathcal{O}_K$ is a Dedekind domain, so any ideal I can be written as $I = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_n^{e_n}$, uniquely up to rearrangement, where the $\mathfrak{p}_i$ are prime ideals of $\mathcal{O}_K$. We will often just refer to these as primes of K. For an extension of number fields L/K, and any prime $\mathfrak{q}$ of L above a prime $\mathfrak{p}$ of K, we get a corresponding extension $\mathbf{F}_\mathfrak{q}/\mathbf{F}_\mathfrak{p}$ of residue fields. We define $f_\mathfrak{q} = [\mathbf{F}_\mathfrak{q} : \mathbf{F}_\mathfrak{p}]$ to be the degree of this extension. Recall that

$$\sum_{\mathfrak{q}|\mathfrak{p}\mathcal{O}_L} e_\mathfrak{q} f_\mathfrak{q} = \deg(L/K).$$

If L/K is a Galois extension of number fields, then its Galois group G acts transitively on the prime ideals of L above a given prime $\mathfrak{p}$ of K. The decomposition group $D_\mathfrak{q}$ for $\mathfrak{q}$ a prime of L is the stabiliser of $\mathfrak{q}$ under this Galois action. Said differently,

$$D_\mathfrak{q} := \{g \in G \ : \ \mathfrak{p}^g = \mathfrak{p}\}\,.$$

Decomposition groups are precisely the Galois groups of the corresponding extensions of local fields, i.e. $D_\mathfrak{q} \cong \mathrm{Gal}(L_\mathfrak{q}/K_\mathfrak{p})$. Decomposition groups are of crucial importance in elementary algebraic number theory. The natural map

$$D_\mathfrak{q} \longrightarrow \mathrm{Gal}(\mathcal{O}_L/\mathfrak{q} \ / \ \mathcal{O}_K/\mathfrak{p})$$

is a surjection with kernel equal to the ramification subgroup $I_\mathfrak{q}$. There is therefore a canonical coset of $I_\mathfrak{q}$ corresponding to the canonical generator $x \mapsto x^{|\mathcal{O}_K/\mathfrak{p}|}$, and when $\mathfrak{q}$ is unramified, this canonical coset consists of a single distinguished element $\mathrm{Frob}_{\mathfrak{q}}$. One of the most powerful and important results in algebraic number theory is the following theorem, which describes the statistical behaviour of these Frobenius elements.

Theorem 27.1 (Chebotarev density). *Let $F/\mathbf{Q}$ be a finite Galois extension, and $\mathcal{C}$ a conjugacy class of $G = \mathrm{Gal}(F/\mathbf{Q})$. Then the set $S_{\mathcal{C}} = \{p \text{ unramified in } F\ : \mathrm{Frob}_p \in \mathcal{C}\}$ has Dirichlet density $|\mathcal{C}|/|G|$.*

For example, if $f \in \mathbf{Z}[x]$ is a monic irreducible quintic with Galois group S_5, then the density of primes that split completely in the Galois closure of f is $1/120$. The density of primes that are inert in this extension is $1/5$. The proof of this theorem is deep, and relies on class field theory. The crucial ingredient is the analytic continuation of certain Artin L-functions. Let us quickly review what these are. Let F/K be a Galois extension of number fields, and ρ a finite dimensional representation of its Galois group. For any prime $\mathfrak{p}$ of K, choose a prime $\mathfrak{q}$ above $\mathfrak{p}$ in F and choose an element $\mathrm{Frob}_{\mathfrak{q}} \in D_{\mathfrak{q}}$, which maps to $(x \mapsto x^{\mathrm{Nm}\mathfrak{q}})$ in the quotient $D_{\mathfrak{q}}/I_{\mathfrak{q}}$. Denote $\mathrm{Frob}_{\mathfrak{p}}$ for the conjugacy class of the element $\mathrm{Frob}_{\mathfrak{q}}$ just defined. The Artin L-function of ρ is defined by the Euler product

$$L(\rho, s) = \prod_{\mathfrak{p}} \frac{1}{\det\left(1 - \mathrm{Nm}(\mathfrak{p})^{-s} \cdot \mathrm{Frob}_{\mathfrak{p}} \mid \rho^{I_{\mathfrak{q}}}\right)},$$

where $\rho^{I_{\mathfrak{q}}}$ is the representation obtained from ρ by restricting to the inertia-invariant vectors. We can prove that this definition is independent of the choices of $\mathfrak{q}$, and this L-series is analytic for $\mathrm{Re}(s) > 1$. One of the holy grails in number theory is to prove the conjecture that whenever ρ is a non-trivial irreducible representation, then its Artin L-series has an analytic continuation to all of $\mathbf{C}$. Hecke proved that when ρ is 1-dimensional, this conjecture holds. The case where ρ is 2-dimensional is now also known thanks to the work of many people in recent decades on modularity of elliptic curves and p-adic families of modular forms.

Artin L-functions satisfy a beautiful *Artin formalism*, of which the most important rules are

- $L(\rho_1 \oplus \rho_2, s) = L(\rho_1, s) \cdot L(\rho_2, s)$,
- If $N \trianglelefteq \mathrm{Gal}(F/K)$ lies in the kernel of ρ, then $L(\rho, s) = L(\rho_N, s)$ where ρ_N is the corresponding representation of $\mathrm{Gal}(F/K)/N \simeq \mathrm{Gal}(F^N/K)$.
- If $\rho = \mathrm{Ind}_H^{\mathrm{Gal}(F/K)} \rho_H$ then $L(\rho, s) \simeq L(\rho_H, s)$.

The proof of all of these properties is entirely formal and straightforward. It turns out that this is already enough to deduce Chebotarev density from Hecke's result on the analytic continuation of Artin L-functions attached to characters. Indeed, induction theorems together with Artin formalism allow us to express $L(\rho, s)$ as a quotient of Artin L-functions for representations induced from characters, which may be used to show that if $\rho \neq \mathbf{1}$, the Artin L-function $L(\rho, s)$ has analytic continuation to $s = 1$ and does not vanish there. Chebotarev density may be proved as a consequence of Hecke's theorem as follows. First, note that if χ_ρ is the character of ρ, we have

$$L(\rho, s) = C(s) \cdot \exp\left(\sum_{\mathfrak{p} \text{ unram}} \sum_{m=1}^{\infty} \chi_\rho(\mathrm{Frob}_p^n) p^{-ns} \right),$$

where C is the product of the ramified factors, which is holomorphic and non-vanishing at $s = 1$. This expression is interesting in its own right, as it shows that Artin L-function only depends on the character χ_ρ in a very explicit way. This shows that

$$\log L(\rho, s) \sim_{s=1} \sum_{\mathfrak{p} \text{ unram}} \chi_\rho(\mathrm{Frob}_p) p^{-s},$$

as the higher order sums are all bounded at $s = 1$. The Chebotarev density theorem then follows from the following sequence of equalities:

$$\begin{aligned} \sum_{p \in S_{\mathcal{C}}} p^{-s} = \sum_{p \text{ unram}} \mathbf{1}_{\mathcal{C}} \cdot p^{-s} &= \sum_\rho \langle \chi_\rho, \mathbf{1}_{\mathcal{C}} \rangle \cdot f_\rho(s) \\ &= \frac{|\mathcal{C}|}{|\mathrm{Gal}(F/\mathbf{Q})|} f_{\mathbf{1}}(s) + \sum_{\rho \neq \mathbf{1}} \langle \chi_\rho, \mathbf{1}_{\mathcal{C}} \rangle f_\rho(s). \end{aligned}$$

Here, $f_\rho(s) = \sum_{\mathfrak{p} \text{ unram}} \chi_\rho(\mathrm{Frob}_{\mathfrak{p}}) p^{-s}$ which is bounded if $\rho \neq \mathbf{1}$ and $\sim \log(s-1)^{-1}$ if $\rho = \mathbf{1}$. This implies the desired statement about the Dirichlet density of primes in $S_{\mathcal{C}}$. So we see that the analytic continuation of an Artin L-series has incredibly powerful consequences.

This is a great result and gives us statistical expectations for the splitting behaviour of a rational prime in a Galois extension. But for any given prime, it does not tell us anything! So how do we experiment with decomposition groups, ramification, and Frobenius elements on concrete examples? The following theorem describes a method to compute the splitting of almost all prime ideals explicitly.

Theorem 27.2 (Kummer–Dedekind). *Let L/K be an extension of number fields. Suppose $\alpha \in \mathcal{O}_L$ with minimal polynomial $f_\alpha \in \mathcal{O}_K[x]$ is such that $N = [\mathcal{O}_L : \mathcal{O}_K[\alpha]]$ is finite. Let $\mathfrak{p} \trianglelefteq \mathcal{O}_K$ be a prime not dividing N, and suppose*

$$f_\alpha(x) \equiv \prod_{i=1}^m g_i(x)^{e_i} \pmod{\mathfrak{p}},$$

where the $g_i \in \mathcal{O}_K[x]$ are distinct and irreducible modulo $\mathfrak{p}$. Then

$$\mathfrak{p}\mathcal{O}_L = \prod_{i=1}^m \mathfrak{q}_i^{e_i}, \qquad \text{where } \mathfrak{q}_i = \mathfrak{p}\mathcal{O}_L + g_i(\alpha)\mathcal{O}_L.$$

The $\mathfrak{q}_i$ are distinct primes of $\mathcal{O}_L$ with ramification index e_i and residue degree equal to $\deg \overline{g_i}(x)$.

This is an extremely practical theorem. The only slightly tedious detail is determining whether for some chosen α a given prime $\mathfrak{p}$ divides the index N. The following criterion is frequently helpful when $K = \mathbf{Q}$:

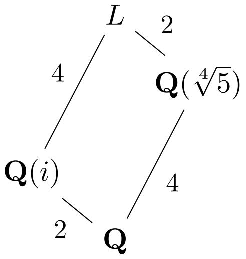
Lemma 27.1. *If $L/\mathbf{Q}$ is a finite extension, $\alpha \in \mathcal{O}_L$ a generator of L over $\mathbf{Q}$ with minimal polynomial $f_\alpha \in \mathbf{Z}[x]$. If $f_\alpha \bmod p$ has distinct roots in $\overline{\mathbf{F}}_p$ then $[\mathcal{O}_L : \mathbf{Z}[\alpha]]$ is coprime to p.*

On any given example, we can now compute lots of examples of prime ideal decomposition, and I strongly urge you to do this if this material is new to you. Here are some examples to get you going:

Example. Let $K = \mathbf{Q}(\sqrt[4]{5})$, and L its Galois closure, which is a D_4-extension of $\mathbf{Q}$. It follows from the above lemma that if $p \neq 2, 5$ then $p \nmid \left[\mathcal{O}_K : \mathbf{Z}[\sqrt[4]{5}]\right]$, as the derivative of its minimal polynomial is $4x^3$. Let us consider the prime $p = 19$ of $\mathbf{Q}$, then we compute the factorisation

$$x^4 - 5 \equiv (x+3)(x+16)(x^2+9) \pmod{19}.$$

It follows from Kummer–Dedekind that $19\mathcal{O}_K = (19, \sqrt[4]{5}+3)(19, \sqrt[4]{5}+16)(19, \sqrt{5}+9)$. This allows us to determine the splitting behaviour of p in L as follows. Because of our calculation in K, there must be at least 3 primes above 19 in L, all of whom have residual degree $f \geq 2$. Since the sum of all of them must be equal to the degree of $L/\mathbf{Q}$, which is 8, there must necessarily be 4 primes in L dividing 19, all have $f = 2$. ■



Example. Let $D \in \mathbf{Z}$ squarefree, and $K := \mathbf{Q}(\sqrt{D})$. It is well known that the complete description of the ring of integers in K depends on the parity of D, and is given by

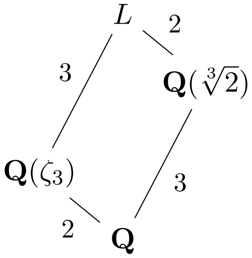
$$\mathcal{O}_K = \begin{cases} \mathbf{Z}[\sqrt{D}] & \text{if } D \equiv 2, 3 \pmod 4 \\ \\ \mathbf{Z}\left[\frac{1+\sqrt{D}}{2}\right] & \text{if } D \equiv 1 \pmod 4 \end{cases}$$

It follows from Kummer-Dedekind that odd primes p split in K if and only if $\left(\frac{D}{p}\right) = 1$. The prime 2 is unramified if and only if $D \equiv 1 \pmod 4$. For instance, pick $D = -6$, $K = \mathbf{Q}(\sqrt{-6})$, then $\mathcal{O}_K = \mathbf{Z}\left[\sqrt{-6}\right]$. By K-D, a prime p is split if and only if $x^2 + 6$ splits mod p. Quadratic reciprocity now lets us conclude that p splits completely if and only if $p \equiv 1, 5, 7, 11 \pmod{24}$. ■

Example. Let us investigate splitting behaviour in the splitting field of $f(x) = x^3 - 2$, which has Galois group S_3 over $\mathbf{Q}$. Take a prime p in $\mathbf{Q}$, then $p \neq 3$ is split in $\mathbf{Q}(\zeta_3)$ if and only if $p \equiv 1 \pmod 3$, and it is inert otherwise. Now let us investigate its splitting behaviour in $K = \mathbf{Q}(\sqrt[3]{2})$. If $p \equiv 2 \pmod 3$, then the map

$$\mathbf{F}_p^\times \to \mathbf{F}_p^\times : x \mapsto x^3$$

is an isomorphism, and so p splits into two primes in K, one of residual degree 1, and one of residual degree 2, by Kummer–Dedekind. If $p \equiv 1 \pmod 3$, the situation is much more subtle. The primes that split completely amongst the primes that are $1 \pmod 3$ are highlighted below:



$$\{7, 13, 19, \mathbf{31}, 37, \mathbf{43}, 61, 67, 73, 79, 97, 103, \mathbf{109}, \ldots\}.$$

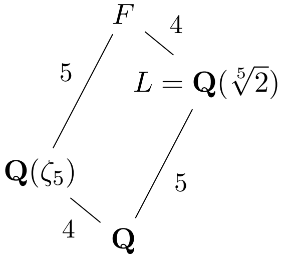
Euler recognised these primes as the ones of the form $x^2 + 27y^2$! We now have a more appealing restatement of Euler's conjecture: A prime $p \neq 3$ is of the form $x^2 + 27y^2$ if and only if it splits completely in L. ■

Of course, it remains mysterious to say anything about splitting behaviour of general primes, and not just particular choices that we manage to do by hand as in the above examples. If we bring in the Galois action on the set of primes, we gain a more complete picture. If F/K is a Galois extension of number fields, then we know that primes in K are in bijection with $\mathrm{Gal}(F/K)$-orbits of primes of F via $\mathfrak{p} \mapsto \{\text{primes above } \mathfrak{p}\}$. The stabiliser of a prime $\mathfrak{q}$ is the decomposition group $D_{\mathfrak{q}}$. The following result describes prime decomposition in intermediate extensions, and will be needed in the sequel.

Theorem 27.3. *Let F/K be a Galois extension of number fields, L an intermediate field, and $\mathfrak{p}$ a prime of K. Then there is a bijection*

$$\{\textit{primes above } \mathfrak{p} \textit{ in } L\} \longleftrightarrow \mathrm{Gal}(F/L)\backslash\mathrm{Gal}(F/K)/D_{\mathfrak{q}},$$

defined by choosing a prime $\mathfrak{q}$ above $\mathfrak{p}$ in F, and sending a prime $\mathfrak{s}$ above $\mathfrak{p}$ in L to the elements that send $\mathfrak{q}$ to some prime above $\mathfrak{s}$ in F.



Example. Let $K = \mathbf{Q}$ and $F = \mathbf{Q}(\zeta_5, \sqrt[5]{2})$. Consider the prime $p = 73$ in $\mathbf{Q}$. Fix primes $\mathfrak{p}$ above (73) in $\mathbf{Q}(\zeta_5)$, and $\mathfrak{q}$ above $\mathfrak{p}$ in F. We may use exercise 1 to conclude that $\mathfrak{p} = 73\mathbf{Z}[\zeta_5]$ is unique, with Frobenius element $\mathrm{Frob}_{\mathfrak{p}}$ is of order 4, and hence $\mathfrak{p}$ has residue degree $f_{\mathfrak{p}/73} = 4$. Also, $\mathfrak{q}/\mathfrak{p}$ is unramified, as otherwise we would have $5 \mid e_{\mathfrak{q}/73}$. This is not possible, as 73 is unramified in $\mathbf{Q}(\sqrt[5]{2})$ since the minimal polynomial $x^5 - 2$ of $\sqrt[5]{2}$ has distinct roots in $\overline{\mathbf{F}}_{73}$, so its factorisation modulo 73, which is necessarily squarefree, gives us the factorisation of (73) in $\mathbf{Q}(\sqrt[5]{2})$ by Kummer–Dedekind. We conclude that $e_{\mathfrak{q}/73} = 1$ and $f_{\mathfrak{q}/73}$ is 4 or 20. As C_{20} is not a subgroup of S_5 it cannot possibly by 20 and hence $D_{\mathfrak{q}} \simeq \mathrm{C}_4$. Now take $L = \mathbf{Q}(\sqrt[5]{2})$, then $\mathrm{Gal}(F/\mathbf{Q})$ permutes the roots of $x^5 - 2$. Without loss of generality we may assume that $D_{\mathfrak{q}}$ fixes $\sqrt[5]{2}$ and permutes the other roots cyclically. This means that there are 2 primes in L above 73, both unramified, one of residue degree 4 and one of residue degree 1. Can you prove this directly using Kummer–Dedekind? ■

We have seen that the Frobenius element $\mathrm{Frob}_{\mathfrak{q}}$ of a prime $\mathfrak{q}$ above $\mathfrak{p}$ depends on $\mathfrak{p}$ up to conjugacy. It is clear that one specific situation is of particular interest: That where conjugation in the Galois group is trivial, or in other words, where the Frobenius element of a prime above $\mathfrak{p}$ only depends on $\mathfrak{p}$. Our ability to attach to a prime of K a well-defined element of the Galois group of an abelian extension is very powerful, as it encodes something external, namely elements in Galois groups of certain extensions of K, in terms of something internal, namely prime ideals of K. Let us be a bit more precise:

Frobenius. We say an extension F/K is abelian if it is Galois with abelian Galois group. For any abelian extension F/K, and any prime $\mathfrak{p}$ of K which is unramified in F/K we write $\mathrm{Frob}_{\mathfrak{p}}$ for the Frobenius element of any prime $\mathfrak{q}$ of F above $\mathfrak{p}$. Note that this is a well-defined element of $\mathrm{Gal}(F/K)$ as conjugation is trivial. In short, there is a well-defined map

$$\{\mathfrak{p} \text{ primes in } \mathcal{O}_K\} \to \mathrm{Gal}(F/K) \ : \ \mathfrak{p} \mapsto \mathrm{Frob}_{\mathfrak{p}}.$$

This map is central to everything related to class field theory. Later, we will extend this to an appropriate group homomorphism and refer to the above map as the *Artin reciprocity map*. For now, it just attaches an element of the Galois group to primes of the base field.

28 Cyclotomic fields

Let us start experimenting with the easiest class of abelian extensions, and see if we can give a nice characterisation of splitting behaviour of primes and Frobenius elements. A field extension F/K is *cyclotomic* if there is an $n \geq 1$ such that F is contained in $K(\zeta_n)$. If $\mathfrak{p} \trianglelefteq \mathcal{O}_K$ is prime such that $\mathfrak{p} \nmid (n)$ then $\mathfrak{p}$ is unramified in F/K and $\mathrm{Frob}_{\mathfrak{p}}$ is the unique element of $\mathrm{Gal}(F/K)$ that sends ζ_n to $\zeta_n^{\mathrm{Nm}\mathfrak{p}}$.

28.1 Cyclotomic extensions of Q.

If $K = \mathbf{Q}$ we can be extremely explicit about the ramification and splitting behaviour of primes in cyclotomic extensions. Let $F = \mathbf{Q}(\zeta_n)$, then a prime p ramifies in F if and only if $p \mid n$. This can easily be seen from Kummer–Dedekind, by first observing that the greatest common divisor of $x^n - 1$ and its derivative nx^{n-1} in $\mathbf{F}_p[x]$ is always 1 when $p \nmid n$. Recall that Frob_p is the unique element in $\mathrm{Gal}(F/\mathbf{Q}) \simeq (\mathbf{Z}/n\mathbf{Z})^\times$ such that

$$\mathrm{Frob}_p(\zeta_n) \equiv \zeta_n^p \pmod{p}.$$

As the n-th roots of unity are all distinct modulo p, we see from this that $\mathrm{Frob}_p = [p] \in (\mathbf{Z}/n\mathbf{Z})^\times$. This leaves nothing to the imagination when it comes to the splitting behaviour of primes in $\mathbf{Q}(\zeta_n)$, see exercise 1.

Remark. Note that Chebotarev density applied to this particular case yields Dirichlet's theorem on primes in arithmetic progressions. It gives us the effective version of Dirichlet's theorem that states that the Dirichlet density of the set of primes in a certain residue class modulo n is precisely $1/\varphi(n)$.

We can just as easily describe the splitting behaviour of primes in intermediate cyclotomic extensions. More precisely, from the above observations it is easy to prove the following theorem:

Theorem 28.1. *Let H be a subgroup of $(\mathbf{Z}/n\mathbf{Z})^\times$, and $F = \mathbf{Q}(\zeta_n)^H$ be the corresponding cyclotomic extension. Considering only primes that are coprime to n, the following are true:*

- *A prime p splits completely in F if and only if $p \in H$.*

- *The map*

$$\phi : \left\{ \frac{a}{b} \in \mathbf{Q} \ : \ a, b \in \mathbf{Z}_{>0},\ \gcd(ab, n) = 1 \right\} \longrightarrow \mathrm{Gal}(F/\mathbf{Q}) \simeq (\mathbf{Z}/n\mathbf{Z})^\times, \ \prod_i p_i^{n_i} \mapsto \prod_i \mathrm{Frob}_{p_i}^{n_i}$$

is a surjective group homomorphism with kernel $\left\{ \frac{a}{b} \pmod{n} \in H \right\}$.

Example. Take $n = 25$ so that $(\mathbf{Z}/n\mathbf{Z})^\times \simeq \mathrm{C}_5 \times \mathrm{C}_4$. Choose H to be the subgroup of squares, which is the unique subgroup isomorphic to C_{10}, so that $F = \mathbf{Q}(\zeta_{25})^H = \mathbf{Q}(\sqrt{5})$. So we see that for a prime p to split completely in F, it is necessary and sufficient that $p \equiv \pm 1 \pmod{5}$. ■

Actually, we can get an interesting result by considerations similar to this example. Consider the *quadratic Gauss sum*

$$G = \sum_{i=0}^{p-1} \left(\frac{i}{p} \right) \zeta_p^i,$$

which is clearly an element of $\mathbf{Q}(\zeta_p)$. One can show by an explicit computation that $G^2 = \left(\frac{-1}{p} \right) p$, from which we see that $F = \mathbf{Q}(\sqrt{\pm p})$ is the unique quadratic subfield of $\mathbf{Q}(\zeta_p)$, where the sign is $+$ when $p \equiv 1 \bmod 4$ and $-$ otherwise. The field F is the fixed field of the unique subgroup of index 2 in $(\mathbf{Z}/p\mathbf{Z})^\times$, which is the subgroup of squares. The above theorem now tells us that a prime $q \neq p$ is a square modulo p if and only if it splits completely in F. But the order $\mathbf{Z}[\sqrt{\pm p}]$ is an order in $\mathcal{O}_F$ of index at most 2, so by Kummer–Dedekind we find that a prime q splits in F if and only if the polynomial

$$x^2 \pm p \in \mathbf{F}_q[x]$$

splits, or equivalently, has a root. We conclude that q is a square mod p if and only if $\pm p$ is a square mod q. Behold: yet another proof of quadratic reciprocity!

28.2 More general cyclotomic extensions.

If we want to generalise the above results for cyclotomic extensions of $\mathbf{Q}$ to more general cyclotomic extensions, we immediately run into a fair number of complications. The corresponding statement is as follows.

Theorem 28.2. *Let K be a number field, and $F = K(\zeta_n)$ be the cyclotomic extension for some n. Considering only primes that are coprime to n, the following are true:*

- *If $\mathfrak{p} = (a)$ is a principal prime of K with $a \equiv 1 \pmod{n}$ and $\sigma(a) > 0$ for every real embedding $\sigma : K \hookrightarrow \mathbf{R}$, then $\mathfrak{p}$ splits completely in F/K.*
- *The map*

$$\phi : \left\{ \frac{a}{b} \in K \ : \ a, b \in \mathcal{O}_K,\ \gcd((ab), (n)) = 1 \right\} \longrightarrow \mathrm{Gal}(F/K) \leq (\mathbf{Z}/n\mathbf{Z})^\times, \ \prod_i \mathfrak{p}_i^{n_i} \mapsto \prod_i \mathrm{Frob}_{\mathfrak{p}_i}^{n_i}$$

is a group homomorphism whose kernel contains

$$P^1_{(n)} = \left\{ (a)(b)^{-1} \; : \; a \equiv b \pmod{n}, \; \sigma(a/b) > 0, \; {}^{\forall}\sigma : K \hookrightarrow \mathbf{R} \right\}.$$

Note that this result is far from being as complete as the theorem for cyclotomic extensions of $\mathbf{Q}$. How do we make it equally precise? Can we get a theorem for all finite abelian extensions, and not just cyclotomic ones? The answers are given by class field theory. We will briefly discuss the statements of global class field theory, and give some pointers about applying these statements in practice. First, we will try to get an idea of what to expect by working out some explicit examples.

Example. Let $K = \mathbf{Q}(i)$ and $F = K(\zeta_3)$. Take $t \in \mathbf{Z}[i]$ and calculate

$$\begin{array}{lclllcl}
\mathrm{Nm}(1+3t) & \equiv & 1 \pmod 3 & & \mathrm{Nm}(i+3t) & \equiv & 1 \pmod 3 \\
\mathrm{Nm}(2+3t) & \equiv & 1 \pmod 3 & & \mathrm{Nm}(2i+3t) & \equiv & 1 \pmod 3 \\
\mathrm{Nm}(1+i+3t) & \equiv & 2 \pmod 3 & & \mathrm{Nm}(2+i+3t) & \equiv & 2 \pmod 3 \\
\mathrm{Nm}(1+2i+3t) & \equiv & 2 \pmod 3 & & \mathrm{Nm}(2+2i+3t) & \equiv & 2 \pmod 3
\end{array}$$

We conclude that if $\mathfrak{p} = (a)$ with $a \equiv \pm 1, \pm i \pmod 3$ then $\mathrm{Frob}_{\mathfrak{p}} = 1$, and if $\mathfrak{p} = (a)$ with $a \equiv \pm 1 \pm i \pmod 3$ then $\mathrm{Frob}_{\mathfrak{p}} : \zeta_3 \mapsto \zeta_3^{-1}$ is conjugation. In this situation, we get behaviour similar to the splitting behaviour in cyclotomic extensions of $\mathbf{Q}$: The Frobenius element is determined by a set of congruences modulo 3. In general, the story is not this simple, as we will see in the next examples. ■

Example. Let $K = \mathbf{Q}(\sqrt{-5})$ and $F = K(\zeta_3)$. We have $\mathcal{O}_K = \mathbf{Z}[\sqrt{-5}]$, and $\mathrm{Cl}_K \simeq \mathrm{C}_2$. By Kummer–Dedekind (3) splits in $\mathcal{O}_K$ and we have $(\mathcal{O}_K/(3))^{\times} \simeq \mathrm{C}_2 \times \mathrm{C}_2 = \langle -1, \sqrt{-5} \rangle$. Take $t \in \mathbf{Z}[\sqrt{-5}]$ and calculate

$$\begin{array}{lcl}
\mathrm{Nm}(\pm 1 + 3t) & \equiv & 1 \pmod 3 \\
\mathrm{Nm}(\pm\sqrt{-5} + 3t) & \equiv & 2 \pmod 3
\end{array}$$

We conclude that if $\mathfrak{p} = (a)$ with $a \equiv \pm 1 \pmod 3$ then $\mathrm{Frob}_{\mathfrak{p}} = 1$, and if $\mathfrak{p} = (a)$ with $a \equiv \pm\sqrt{-5} \pmod 3$ then $\mathrm{Frob}_{\mathfrak{p}} : \zeta_3 \mapsto \zeta_3^{-1}$ is conjugation. But there are non-principal ideals in $\mathcal{O}_K$, so this is not the end of the story! Take $\mathfrak{p}_2 = (2, 1+\sqrt{-5})$ the prime above 2. We have $\mathfrak{p}_2^2 = (2)$ so it follows that $\mathrm{Nm}\mathfrak{p}_2 = 2$. We conclude that $\mathrm{Frob}_{\mathfrak{p}_2} : \zeta_3 \mapsto \zeta_3^2 = \zeta_3^{-1}$ is conjugation. In fact, this is enough to determine Frobenius for all non-principal primes. Indeed, for any non-principal prime $\mathfrak{p}$ we have that $\mathfrak{p} \cdot \mathfrak{p}_2 = (a)$ is principal, and so

$$\mathrm{Frob}_{\mathfrak{p}} = \begin{cases} 1 & \text{if} \quad a \equiv \pm\sqrt{-5} \pmod 3 \\ \zeta_3 \mapsto \zeta_3^{-1} & \text{if} \quad a \equiv \pm 1 \pmod 3 \end{cases}$$

This shows that Frobenius depends on the image of $\mathfrak{p}$ in the group of ideals modulo "principal ideals $\equiv 1 \pmod 3$", which is a combination of the class of $\mathfrak{p}$ in the ideal class group, and a congruence modulo 3. The next example shows that the general answer can be even more complicated. ■

Example. Let $K = \mathbf{Q}(\sqrt{3})$ and $F = K(\zeta_3)$. We have $\mathcal{O}_K = \mathbf{Z}[\sqrt{3}]$, and $\mathrm{Cl}_K = 1$. By Kummer–Dedekind, we have that 3 is ramified in $\mathbf{Z}[\sqrt{3}]$. Now set $\mathfrak{p} = (1 + \sqrt{3})$, and calculate $\mathrm{Nm}\ \mathfrak{p} = |\mathrm{Nm}(1 + \sqrt{3})| = 2$, so that

$$\mathrm{Frob}_{\mathfrak{p}} : \zeta_3 \mapsto \zeta_3^{-1}.$$

On the other hand, the prime ideal $\mathfrak{q} = (4+\sqrt{3})$ gives us $\mathrm{Nm}\ \mathfrak{q} = |\mathrm{Nm}(4+\sqrt{3})| = 13 \equiv 1 \pmod 3$, yielding instead

$$\mathrm{Frob}_{\mathfrak{q}} : \zeta_3 \mapsto \zeta_3.$$

It follows that unramified primes in the same ideal class (note that the class group is trivial!) can give rise to different Frobenius elements, even if their generators are congruent modulo 3. In this case, Frobenius is determined by the image of $\mathfrak{p}$ in the group of ideals modulo "principal ideals with totally positive generator". This is reminiscent of the definition of narrow class groups, and it is no accident that we chose a quadratic field K for which $\mathrm{Cl}_K = 1$ but $\mathrm{Cl}_K^+ = 2$. ■

29 Global class field theory: Results

Everything is most concisely expressed in the language of adèles. Given a number field K, its ring of adèles is defined as

$$\mathbf{A}_K = \prod_v{}' K_v$$

where the product runs over all places of K, and the notation $\prod'$ means that $(a_v)_v \in \mathbf{A}_K$ must have $a_v \in \mathcal{O}_v$ for all but finitely many v. We can topologise $\mathbf{A}_K$ by setting $\prod_v \mathcal{O}_v$ with its product topology to be open. There is a diagonal map

$$\Delta : K \hookrightarrow \mathbf{A}_K,$$

which endows K with the discrete topology. The quotient $\mathbf{A}_K/K$ is compact. The units in $\mathbf{A}_K$ form a group with respect to multiplication, which we will call the *idèle group*. We topologise it, not with the subspace topology from $\mathbf{A}_K$, but simply by declaring $\prod_v \mathcal{O}_v^\times$, with its product topology, to be open in $\mathbf{A}_K^\times$. The image of $K^\times$ under the diagonal map is again discrete, and the (non-compact) quotient

$$C_K = \mathbf{A}_K^\times / K^\times$$

is called the *idèle class group* of K. It plays the lead role in global class field theory.

Fix a separable closure of K, and take the maximal abelian subextension K^{ab}/K. Then there exists a *global Artin map*

$$\varphi : C_K \to \mathrm{Gal}(K^{\mathrm{ab}}/K).$$

This map is surjective, and its kernel is the connected component of the identity. It becomes an isomorphism of topological groups when we pass to the *profinite completion*. More precisely, we have

$$\varphi : \widehat{C}_K \xrightarrow{\sim} \mathrm{Gal}(K^{\mathrm{ab}}/K), \qquad \text{where } \widehat{C}_K = \varprojlim_U C_K/U,$$

with the limit taken over all finite index open subgroups. This map is compatible with base change in the sense that whenever we consider a finite extension L/K we have that $\mathbf{A}_L \simeq \mathbf{A}_K \otimes L$ and hence there is a well-defined norm map

$$\mathrm{Nm}_{L/K} : \mathbf{A}_L \longrightarrow \mathbf{A}_K,$$

which induces a corresponding norm map on idèle class groups. The following diagram now commutes:

$$\begin{tikzcd}
C_L \arrow[r, "\varphi_L"] \arrow[d, "\mathrm{Nm}_{L/K}"'] & \mathrm{Gal}(L^{\mathrm{ab}}/L) \arrow[d, "\mathrm{Res}"] \\
C_K \arrow[r, "\varphi_K"] & \mathrm{Gal}(K^{\mathrm{ab}}/K)
\end{tikzcd}$$

We will usually be interested in all finite abelian extensions. All the information on these is contained in the topological properties of the Artin map. Indeed, the Artin map sets up a bijection between finite index open subgroups of C_K, and finite index open subgroups of $\mathrm{Gal}(K^{\mathrm{ab}}/K)$, which by Galois theory correspond to finite extensions of K. We can be even more specific by the above compatibility with field extensions: A finite abelian extension L/K corresponds to the finite index open subgroup $\mathrm{Nm}_{L/K}C_L$ of C_K. This is a powerful dictionary, as it encodes information about extensions of K into an object that makes no reference to other fields and depends only on the internal arithmetic of K. Herein lies the power of class field theory.

30 Global class field theory: Practice

The above summary of global class field theory is very concise, and leaves us with little confidence in practice. Clearly, the dictionary between finite abelian extensions of a number field K (inaccessible) and finite index subgroups of the idèle class group (accessible) is the key interest, so we first ask ourselves whether we can make such finite index open subgroups explicit.

A modulus is a formal product $\mathfrak{m} = \mathfrak{m}_f \mathfrak{m}^f = \prod_v v^{e_v}$ where v runs over all places of K, such that

- $e_v \in \mathbf{Z}_{\geq 0}$ and $e_v = 0$ for almost all v.
- When v is real, $e_v \in \{0, 1\}$, when v is complex, $e_v = 0$.

Clearly, we can (and will) think of a modulus as a pair consisting of an ideal in $\mathcal{O}_K$ and a subset of the set of real palces of K. Given a modulus, we associate $U_{\mathfrak{m}} \subseteq C_K$ which is defined to be the image of $\prod_v U_{\mathfrak{m},v} \subset \mathbf{A}_K^\times$ where $U_{\mathfrak{m},v} = \mathcal{O}_v^\times$ if $e_v = 0$, and otherwise:

- v finite: $U_{\mathfrak{m},v} = 1 + \mathfrak{p}_v^{e_v} \subseteq \mathcal{O}_v^\times$
- v real: $U_{\mathfrak{m},v} = \mathbf{R}_{>0}^\times$

Then $U_\mathfrak{m}$ is a finite index open subgroup of C_K, and moreover any finite index open subgroup of C_K contains $U_\mathfrak{m}$ for some modulus $\mathfrak{m}$. The finite extension of K corresponding to $U_\mathfrak{m}$ is called the *ray class field* $K_\mathfrak{m}$ attached to $\mathfrak{m}$. Its Galois group $C_K/U_\mathfrak{m}$ over K is called the *ray class group* $\mathrm{Cl}_\mathfrak{m}$ of modulus $\mathfrak{m}$.

We can be very explicit about ray class groups. Let us introduce two pieces of notation:

$$I_\mathfrak{m} = \bigoplus_{\mathfrak{p} \nmid \mathfrak{m}} \mathfrak{p}^\mathbf{Z}, \qquad P_\mathfrak{m}^1 = \left\{ (a) \ \text{ for } a \equiv 1 (\operatorname{mod}^\times \mathfrak{m}) \right\},$$

where $a \equiv 1(\operatorname{mod}^\times \mathfrak{m})$ means that $a \in U_{\mathfrak{m},v}$ for all v such that $e_v > 0$. In other words, $I_\mathfrak{m}$ is the group of fractional ideals coprime to $\mathfrak{m}$, and $P_\mathfrak{m}^1$ is the group of principal ideals generated by those elements which are 1 mod $\mathfrak{m}$ and positive under every embedding in $\mathfrak{m}$. Finally, the ray class group is now easily seen to be isomorphic to $I_\mathfrak{m}/P_\mathfrak{m}^1$. Clearly, for $\mathfrak{m} = (1)$ we obtain the usual class group Cl_K of the ring of integers in K. The associated ray class field is commonly called the *Hilbert class field* of K.

It turns out that any finite open subgroup of C_K contains $U_\mathfrak{m}$ for some modulus $\mathfrak{m}$. Therefore any finite abelian extension of K is contained in the ray class field $K_\mathfrak{m}$ of some modulus $\mathfrak{m}$. A *congruence subgroup* H for $\mathfrak{m}$ is a subgroup of $I_\mathfrak{m}$ containing $P_\mathfrak{m}^1$. An extension F/K is called a *class field* for $(\mathfrak{m}, H)$ if the primes $\mathfrak{p} \nmid \mathfrak{m}$ of K that split completely in F are precisely those that lie in H. We obtain the following more down-to-earth reformulation of the main theorems of class field theory.

Theorem 30.1 (Takagi, Artin). *Let K be a number field. Then the following hold:*

- *Every $(\mathfrak{m}, H)$ has a unique class field F, which is abelian over K.*
- *Every abelian extension F/K is a class field for some $(\mathfrak{m}, H)$. There is a unique minimal such $(\mathfrak{m}, H)$ called the conductor of F/K. The primes that ramify in F/K are precisely those that divide the conductor.*
- *(Artin reciprocity) If F/K is a class field for $(\mathfrak{m}, H)$, then the Artin map*

$$\varphi : \mathrm{Cl}_\mathfrak{m} := I_\mathfrak{m}/P_\mathfrak{m}^1 \to \mathrm{Gal}(F/K) \ : \ \mathfrak{p} \mapsto \mathrm{Frob}_\mathfrak{p},$$

is a surjective homomorphism with kernel H.

This is a much more explicit version that will suit our needs very well. Before we proceed to make things even more explicit, let us point out that by setting $K = \mathbf{Q}$ we see that every modulus is of the form $\mathfrak{m} = (N)$ or $\mathfrak{m} = (N)\infty$. Cyclotomic fields are class fields for all $(N)\infty$, and hence by uniqueness exhaust all abelian extensions. We get the following celebrated theorem.

Corollary 30.1 (Kronecker–Weber). *All abelian extensions of $\mathbf{Q}$ are cyclotomic.*

Now, let us do some explicit computations. First, we will want to compute information about $\mathrm{Cl}_{\mathfrak{m}}(\mathcal{O}_K)$. Let $P_{\mathfrak{m}_f}$ be the subgroup of $I_{\mathfrak{m}}$ generated by the principal ideals prime to $\mathfrak{m}_f$, and $P_{\mathfrak{m}}$ the subgroup of $P_{\mathfrak{m}_f}$ generated by those elements that are positive under every real embedding of $\mathfrak{m}$. Then we can see $\mathrm{Cl}_{\mathfrak{m}}$ as a successive quotient of groups:

$$\begin{aligned} I_{\mathfrak{m}}/P_{\mathfrak{m}_f} &= I_{\mathfrak{m}}/P_1 \cap I_{\mathfrak{m}} \simeq I_{\mathfrak{m}}P_1/P_1 \simeq \mathrm{Cl}_K, \\ P_{\mathfrak{m}_f}/P_{\mathfrak{m}} &\leq \mathrm{C}_2^{\#\mathfrak{m}^f}, \\ P_{\mathfrak{m}}/P_{\mathfrak{m}}^1 &\simeq (\mathcal{O}_K/\mathfrak{m}_f)^\times/\mathcal{O}_K^\times. \end{aligned}$$

This gives us a lot of information, which, given all groups in sight are abelian, is often enough to determine the ray class group up to isomorphism.

Example. Set $K = \mathbf{Q}(i)$ and take modulus $\mathfrak{m} = (1)$. Note that $\mathcal{O}_K = \mathbf{Z}[i]$ is a UFD, which shows that $\mathrm{Cl}_{\mathfrak{m}} = 1$ and so K is its own Hilbert class field. ■

Example. Set $K = \mathbf{Q}(i)$ and take modulus $\mathfrak{m} = (3)$. Note that $\mathcal{O}_K = \mathbf{Z}[i]$ is a UFD, which simplifies everything. Two of the three factors in our successive quotients above are trivial, leaving us with

$$\mathrm{Cl}_{\mathfrak{m}} \simeq (\mathbf{Z}[i]/(3))^\times/\{\pm 1, \pm i\} \simeq \mathrm{C}_2.$$

This implies that the ray class field $K_{\mathfrak{m}}$ is a quadratic extension unramified outside 3, such that all the primes that split in it are those congruent to 1 mod 3. The cyclotomic extension $K(\zeta_3)$ satisfies these conditions, and must therefore be the ray class field associated to $\mathfrak{m}$ by uniqueness. ■

In general, it is extremely difficult to explicitly find class fields attached to a certain modulus, and we were lucky in all the examples we have done so far. In very exceptional situations, we know a lot. For instance, over $\mathbf{Q}$ we know by Kronecker–Weber that all abelian extensions are cyclotomic, and we can hence explicitly construct those extensions by adjoining appropriate combinations of roots of unity. Another particular case is that of imaginary quadratic fields, where we can obtain abelian extensions by adjoining values of the j-function, which is precisely the goal of the second chapter of these notes. Generalising these explicit descriptions to other number fields is commonly known as the *Kronecker Jugendtraum* and is the subject of Hilbert's 12th problem. Even the setting of real quadratic fields is mysterious, though there is a lot of fascinating recent work that remains largely conjectural. Sometimes, we can make due with some versatile basic tools, such as Kummer theory: If K is a number field that contains ζ_n for some n, then every C_n-extension of K is of the form $K(\sqrt[n]{a})$ for some $a \in K$. Indeed, this follows from the statement

$$\mathrm{H}^1(K, \mu_n) \simeq K^\times/(K^\times)^n$$

which is mentioned as a consequence of Hilbert 90 in Appendix B. To see how this is useful, we will illustrate this on an example.

Example. Let $K = \mathbf{Q}(i)$ and $\mathfrak{m} = (7)$. Then we find that

$$\mathrm{Cl}_{\mathfrak{m}} \simeq (\mathbf{Z}[i]/(7))^\times/\{\pm 1, \pm i\} \simeq \mathrm{C}_{48}/\mathrm{C}_4 \simeq \mathrm{C}_{12}.$$

Take $I_{\mathfrak{m}} \supset H \supset P^1_{\mathfrak{m}}$ with $I_{\mathfrak{m}}/H \simeq \mathrm{C}_4$. The class field of $(\mathfrak{m}, H)$ has Galois group C_4 and so by Kummer theory it is of the form $K(\sqrt[4]{a})$ for some $a \in K$. We may assume without loss of generality that $a \in \mathcal{O}_K$ and that it is "fourth-power free". Any prime $\mathfrak{p}$ that divides a must also ramify in the class field, and 7 is the unique prime that ramifies in it, so $a = i^m 7^n$ for some $0 \leq m, n \leq 3$. We cannot have $n = 0, 2$ for otherwise $K(\sqrt{a})$ would be an unramified extension of K which does not exist as shown in a previous example. Therefore we may assume $n = 1$, and it remains to find m. This can be done by computing the Frobenius element attached to some prime other than 7. Take $(53) = (7+2i)(7-2i)$ and notice that $\mathfrak{p} = (7-2i)$ is contained in H so that $\mathrm{Frob}_{\mathfrak{p}}$ is trivial. This means that

$$\sqrt[4]{a} = \mathrm{Frob}_{\mathfrak{p}}(\sqrt[4]{a}) \equiv \sqrt[4]{a}^{53} \equiv -i^m \sqrt[4]{a} \pmod{\mathfrak{p}}.$$

This implies that $m = 2$ and hence the class field is $\mathbf{Q}(i, \sqrt[4]{-7})$. ■

31 The principal ideal theorem

One the most famous results on Hilbert class fields is the so-called *principal ideal theorem*, which states that every ideal in a number field K becomes principal in the Hilbert class field of K. Apart from a difficult result in group theory due to Furtwängler, its proof uses some beautiful properties of ray class fields, which are very useful by themselves.

Ray class fields. Let K/k be a Galois extension of number fields and $\mathfrak{m}$ a modulus of K with $\mathfrak{m} = \mathfrak{m}^g$ for all $g \in \mathrm{Gal}(K/k)$. Then $K_{\mathfrak{m}}$ is Galois over k. If $\mathfrak{n} \mid \mathfrak{m}$ then $K_{\mathfrak{n}^g} = K_{\mathfrak{n}}^{\tilde{g}}$ for any lift $\tilde{g} \in \mathrm{Gal}(K_{\mathfrak{m}}/k)$ of g. Finally, we have

$$\varphi(\mathfrak{a}^g) = \tilde{g}\varphi(\mathfrak{a})\tilde{g}^{-1}.$$

where $\varphi : I_{\mathfrak{m}} \to \mathrm{Gal}(K_{\mathfrak{m}}/K)$ is the Artin map.

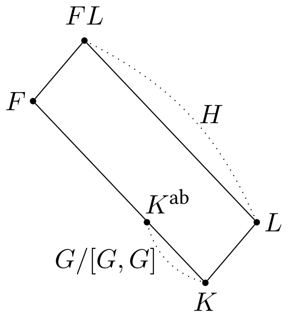
Hilbert class fields. Let K be a number field, and $F = K_{(1)}$ its Hilbert class field. This extension plays a central role in the theory of complex multiplication. We know that $\mathrm{Gal}(F/K) \simeq \mathrm{Cl}_K$, and from the general theory it follows that:

Corollary 31.1. *A prime $\mathfrak{p}$ of K is principal if and only if $\mathfrak{p}$ splits completely in F/K. More precisely, the order of $\mathfrak{p}$ in Cl_K is equal to the residue degree of $\mathfrak{p}$ in F/K. The Hilbert class field F/K is unramified at all primes, and all embeddings $K \hookrightarrow \mathbf{R}$ extend to $F \hookrightarrow \mathbf{R}$. If L/K is any other abelian extension with this property, then we must have $L \subseteq F$.*

The transfer map. Let $H \leq G$ be finite groups. The transfer map, or *Verlagerung*, is the map

$$\mathrm{Ver} : G/[G,G] \longrightarrow H/[H,H]$$

defined by taking right cosets $Hr_1, Hr_2, \ldots, Hr_n$ of H in G, and letting $Hr_i g = Hr_{\sigma(i)}$ for some $g \in G$ and then setting $\mathrm{Ver}(g) = \prod_i r_i g r_{\sigma(i)}^{-1}[H,H]$. This is a well-defined homomorphism, that naturally comes up in the following setting: Consider a Galois extension F/K, with Galois group G.



Now consider a finite extension L/K such that the Galois extension FL/L is abelian with group H. Denoting K^{ab} for the maximal abelian extension of K in F, we obtain the following diagram:

Note that $H \leq G$, as any automorphism of FL/K restricts to an automorphism of F/K, and if it acts trivially on F and L, it must act trivially on FL. This induces a natural map $q : H \to G/[G,G]$ which makes the following diagram commute

$$\begin{tikzcd}
I_{\mathfrak{m}} \arrow[r, "\varphi_{FL/L}"] \arrow[d, "\mathrm{Nm}_{L/K}"'] & H \arrow[d, "q"] \\
I_{\mathrm{Nm}\mathfrak{m}} \arrow[r, "\varphi_{F/K}"] & G/[G,G]
\end{tikzcd}$$

It suffices to check this on prime ideals, as the norm map Nm is multiplicative. Let $\mathfrak{q}$ be a prime of L above $\mathfrak{p}$ unramified, then

$$\varphi_{F/K}(\mathrm{Nm}_{L/K}\mathfrak{q}) = \varphi_{F/K}(\mathfrak{p}^{f_\mathfrak{p}}) = \mathrm{Frob}_\mathfrak{p}^{f_\mathfrak{p}}(F/K) = \mathrm{Frob}_\mathfrak{q}(FL/L) = \varphi(\mathfrak{q}).$$

But what does this have to do with the transfer map? Well, note that we are working in the abelianisation of H, so we may rearrange the product as

$$\mathrm{Ver}(g) = (r_1 g r_{\sigma(1)^{-1}})(r_{\sigma(1)} g r_{\sigma^2(1)^{-1}}) \dots = \prod_{s \in \Sigma} s g^{f(s)} s^{-1},$$

where Σ is a set of double coset representatives $H \backslash G / \langle g \rangle$ and $f(s)$ is the length of the orbit of Hs under $\langle g \rangle$. Now for a prime $\mathfrak{p}$ of K that is unramified, we calculate

$$\begin{aligned}
\mathrm{Ver}\varphi_{F/K}(\mathfrak{p}) = \mathrm{Ver}\mathrm{Frob}_\mathfrak{p} &= \prod_{s \in \Sigma} s \mathrm{Frob}_\mathfrak{p}^{f(s)} s^{-1} \\
&= \prod_{\mathfrak{q} | \mathfrak{p}} \mathrm{Frob}_\mathfrak{q} \\
&= \prod_{\mathfrak{q} | \mathfrak{p}} \varphi_{FL/L}\mathfrak{q} \\
&= \varphi_{FL/L}(\mathfrak{p}\mathcal{O}_L)
\end{aligned}$$

This proves the following result.

Corollary 31.2. *Let F/K be a Galois extension with group G. Let L/K be a finite extension and $\mathfrak{m}$ the modulus of the maximal abelian extension of K in F. Then the following diagram commutes:*

$$\begin{tikzcd}
I_{\mathfrak{m}} \arrow[r, "\varphi_{F/K}"] \arrow[d, "\mathrm{Conorm}_{L/K}"'] & G/[G,G] \arrow[d, "\mathrm{Ver}"] \\
I_{\mathfrak{m}\mathcal{O}_L} \arrow[r, "\varphi_{FL/L}"] & H/[H,H]
\end{tikzcd}$$

Now consider the following special case of the above setting: Set $L = K_{(1)}$ the Hilbert class field of K, and $F = L_{(1)}$ its Hilbert class field. By our discussion on ray class fields above, it follows that F/K is indeed Galois. With notation as above, it follows from our discussion on Hilbert class fields that $H = [G, G]$. If $\mathfrak{p}$ is a prime of K, then we have

$$\varphi_{F/L}(\mathfrak{p}\mathcal{O}_L) = \mathrm{Ver}(\mathrm{Frob}_{\mathfrak{p}}).$$

If we could prove something about the transfer map in this setting, this would restrict the possible splitting behaviour of ideals $\mathfrak{p}\mathcal{O}_L$ in F. In fact, we have the following *principal ideal theorem* conjectured by Hilbert, and proved in 1929 by Furtwängler.

Theorem 31.1 (Artin, Furtwängler). *Let G be a finite group and $H = [G, G]$, then the transfer map is trivial. As a consequence, all ideals in a number field K become principal in its Hilbert class field.*

32 Exercises

1. Let p, q be two distinct prime numbers, and let t be the order of q in $\mathbf{F}_p^\times$. Show that there are $(p-1)/t$ primes in $\mathbf{Q}(\zeta_p)$ above q, each unramified with residue degree t.
2. Let $K = \mathbf{Q}(\zeta_{15})$ determine the factorisations, as well as the decomposition and inertia groups, for the primes $p = 2, 3, 5, 61$.
3. Find explicit generators for all the subfields of the extension $\mathbf{Q}(\zeta_{13})/\mathbf{Q}$.
4. Determine which primes ramify in $\mathbf{Q}(\sqrt[3]{2})$.
5. Find a finite extension of $\mathbf{Q}$ in which no primes are inert.
6. Let F/K be a Galois extension of number fields with a cyclic Galois groups of order p^n for some prime p, and let L/K be the unique subextension of degree p. Show that if a prime $\mathfrak{q}$ of K is inert in L/K and unramified in F, then it is inert in F/K, and that if it is ramified in L/K, then it is totally ramified in F/K.

7. Let $\mathfrak{q} = (11, \zeta_5 - 4)$, which is an unramified prime of $\mathbf{Q}(\zeta_5, \sqrt[5]{5})/\mathbf{Q}$ above 11 of residue degree 5. Determine the decomposition group $D_\mathfrak{q}$ and find $\mathrm{Frob}_\mathfrak{q}$.

8. (Euler's criterion) Let $f = x^3 + bx + c \in \mathbf{Z}[x]$ and define

$$t = \mathrm{Tr}\begin{pmatrix} 0 & 0 & -c \\ 1 & 0 & -b \\ 0 & 1 & 0 \end{pmatrix}^{p+1}.$$

Prove that if $p \nmid b$ then

- f has 3 roots in $\mathbf{F}_p$ if and only if $t \equiv -2b \pmod p$,
- f has 1 root in $\mathbf{F}_p$ if and only if $t^3 - 3b^2t - 2b^3 - 27c^2 \equiv 0 \pmod p$,
- f has 0 roots in $\mathbf{F}_p$ if and only if $t \equiv b \pmod p$,

Hint: Consider fixed points of $\mathrm{Frob}_\mathfrak{q}$ on the roots α, β, γ of f, where $\mathfrak{q}$ is a prime above p in the splitting field of f. What is the minimal polynomial of $\alpha \cdot g(\alpha) + \beta \cdot g(\beta) + \gamma \cdot g(\gamma)$ for $g \in \mathrm{Gal}(f)$?

9. Consider the elliptic curve $X_1(11) : y^2 + y = x^3 - x^2$, which has 5 rational points. Let K be a number field with r_1 real embeddings, and r_2 pairs of complex embeddings, and s_{11} the number of primes above 11 in K. The Birch–Swinnerton-Dyer conjecture (which you may assume for this exercise) predicts that whenever the expression

$$r_1 + r_2 + s_{11}$$

is odd, then there are infinitely many K-rational points on $X_1(11)$. Find all quadratic extensions of $\mathbf{Q}$ for which this criterion is satisfied. Find all number fields of the form $\mathbf{Q}(\sqrt[7]{a})$ for which this criterion is satisfied.

10. Find a number field with two primes above 2, three primes above 3, and five primes above 5.

11. Let $\mathcal{S}$ be a set of primes. Define its natural density to be

$$\lim_{n \to +\infty} \frac{|\{p \in \mathcal{S} \;:\; p \leq n\}|}{|\{p \leq n\}|},$$

provided this limit exists. Show that if the natural density of the set $\mathcal{S}$ exists, then show that its Dirichlet density also exists, and is equal to the natural density.

12. Let ρ be the unique 2-dimensional irreducible representation of $\mathrm{Gal}(\mathbf{Q}(\zeta_3, \sqrt[3]{5})/\mathbf{Q})$. Compute the first ten coefficients of its Artin L-series.

13. Let $f(x) \in \mathbf{Z}[x]$ be a monic irreducible polynomial of degree $d \geq 2$. Consider the set

$$N(f) = \{\text{primes } p \;:\; f(x) \text{ has no roots in } \mathbf{F}_p\}.$$

Show that the Dirichlet density of $N(f)$ is at least $1/d$, and give a sufficient condition for it to be equal to $1/d$. Show that equality is possible, but is not always satisfied, by giving appropriate examples.

14. Let $F/\mathbf{Q}$ be a finite Galois extension, and let ρ and τ be two representations of its Galois group. Prove that if $L(\rho, s) = L(\tau, s)$, then $\rho \simeq \tau$.

15. Let K, L be Galois extensions of $\mathbf{Q}$ inside some fixed closure $\overline{\mathbf{Q}}$. Show that it is true that every prime that splits completely in K, also splits completely in L, if and only if $L \subseteq K$.

16. Let K be a number field, then define its Dedekind zeta function
$$\zeta_K(s) = \sum_{(0) \neq \mathfrak{a} \trianglelefteq \mathcal{O}_K} \mathrm{Nm}_{K/\mathbf{Q}}(\mathfrak{a})^{-s}.$$
Show that if K, L are two finite Galois extensions of $\mathbf{Q}$ with the same Dedekind zeta function, they must be isomorphic.

17. Suppose K is a number field and $\alpha \in \mathcal{O}_K$ is a square modulo all but finitely many prime ideals $\mathfrak{p}$. Prove that α is already a square in $\mathcal{O}_K$.

18. Let $K = \mathbf{Q}(\zeta_3)$ and $\mathfrak{a} = (7)$. Show that there is a unique congruence subgroup H with $I_\mathfrak{a}/H \simeq \mathrm{C}_3$, and describe this subgroup in terms of congruence classes of $\mathbf{Z}[\zeta_3]$ modulo (7). Find the class field of $(\mathfrak{a}, H)$.

19. Find a quadratic extension of $\mathbf{Q}(i)$ that is unramified outside 11, and prove it is unique. Show that there is no such C_4-extension.

20. Prove that for any odd prime p, the unique quadratic subfield of $\mathbf{Q}(\zeta_p)$ has odd class number.

21. Show that if $\mathbf{Q}(\sqrt{n})$ has ideal class group C_4, then its Hilbert class field has Galois group D_4 over $\mathbf{Q}$. Then show that every dihedral group is the Galois group of some extension of $\mathbf{Q}$.

22. Find a number field whose class number is divisible by 11.

23. Prove that the ray class field with modulus $\mathfrak{a} = (41)$ in $\mathbf{Z}[i]$ contains exactly two subfields which are Galois over $\mathbf{Q}$, with Galois groups C_{10} and D_5 respectively.

24. Let $K = \mathbf{Q}(\sqrt{2})$, which has two real embeddings σ_1, σ_2. Compute $I_\mathfrak{a}/P_\mathfrak{a}^1$ for $\mathfrak{a} = (3), \{(3), \sigma_1\}$ and $\{(3), \sigma_1, \sigma_2\}$. For each of these, determine the corresponding ray class field.

25. Let $\mathfrak{m} = (5 + 3\zeta_3)$ be a modulus of $\mathbf{Q}(\zeta_3)$. Find the ray class field of $\mathfrak{m}$.

26. Prove that there are no C_4-extensions of $\mathbf{Q}$ whose quadratic subfield is $\mathbf{Q}(\zeta_3)$.

Bibliography

[Bue89] D. Buell. *Binary Quadratic Forms.* Springer-Verlag, 1989. ↑14, 18.

[Con97] J. H. Conway. *The sensual (quadratic) form.* The Carus mathematical monographs. The Mathematical Association of America, 1997. ↑20, 24.

[Cox89] D. Cox. *Primes of the form* $x^2 + ny^2$. Wiley-Interscience, 1989. ↑6, 9, 14, 70.

[CR99] H. Cohen and X.-F. Roblot. Computing the Hilbert class field of real quadratic fields. *Math. Comp.*, 69(231):1229–1244, 1999. ↑71.

[Fla88] D. Flath. *Introduction to number theory.* Wiley, 1988. ↑31, 33.

[Hat17] A. Hatcher. *The topology of numbers.* Available at `https://www.math.cornell.edu/ hatcher/TN/TNpage.html`, 2017. ↑20, 21.

[Poo] B. Poonen. A brief summary of the main results of class field theory. ↑93.

[Ser70] J.-P. Serre. *Cours d'Arithmétique.* Presses Universitaires de France, 1970. ↑15.

[Ser80] J.-P. Serre. *Trees.* Springer-Verlag, 1980. ↑20.

[Shi70] G. Shimura. On canonical models of arithmetic quotients of bounded symmetric domains. *Ann. of Math.*, 91:144–222, 1970. ↑6, 53, 54, 58, 71.

[Sil09] J. Silverman. *The arithmetic of elliptic curves, 2nd edition*, volume 106 of *GTM*. Springer-Verlag, 2009. ↑6, 53.

[ST68] J.-P. Serre and J. Tate. Good reduction of abelian varieties. *Ann. of Math.*, 68:492–517, 1968. ↑61.

[Wat03] M. Watkins. Class numbers of imaginary quadratic fields. *Math. Comp.*, 73(246):907–938, 2003. ↑16.