The Ramanujan Constant

An Essay on Elliptic Curves, Complex Multiplication and Modular Forms

B.J.Green

Contents

0.1 Introduction 1

1 **3**

1.1 Introduction to Elliptic Curves 3
1.1.1 The Group Law 3
1.1.2 The Endomorphism Ring 4
1.2 Elliptic Curves as Complex Tori 5
1.3 Structure of End(E) 9
1.3.1 Consequences of Theorem **??** 11
1.4 The j-invariant 13
1.4.1 j-invariant of a lattice 14
1.4.2 Surjectivity of j 15
1.4.3 Galois actions on Elliptic Curves 21

2 **23**

2.1 Modular Groups, Curves and Functions 24
2.1.1 $X(1)$ 24
2.1.2 Higher Modular Groups and Curves 26
2.2 The Modular Equation, and the Integrality of the j-Invariant 31

3 **37**

3.1 Modular Forms 37
3.2 q-series for j 41
3.2.1 q-Expansions of Eisenstein Series 41
3.2.2 Jacobi's Product Formula for Δ 42
3.2.3 Properties of $c(n)$ and $\boldsymbol{\tau}(n)$ 47
3.3 Bounds on the coefficients $c(n)$ 48

0.1 Introduction

Ramanujan observed that $e^{\pi\sqrt{163}} = 262537412640768743.99999999999925$ is within 10^{-12} of an integer and used this to obtain approximations to π. In his Field's Medal lecture, Richard Borcherds said that every mathematician should see once in his/her life why this should be the case, and this essay

is an attempt to do just that. However this goal is really just an excuse to study some classic topics in the mathematics of elliptic curves and modular forms. The essay contains three Chapters.

In the first Chapter we discuss the properties of elliptic curves over $\mathbb{C}$, and show that they are in some sense the same thing as lattices in $\mathbb{C}$. In proving this result, the so-called Uniformisation Theorem for Elliptic Curves over $\mathbb{C}$, we find it necessary to introduce the j-invariant and the first modular group $\Gamma(1)$. Both of these will appear time and again in Chapters 2 and 3. It is perhaps worth noting that the entire subject can be approached without mention of Elliptic Curves- however I feel that such an approach does not place the subject in its proper context.

In Chapter 2 we look at $\Gamma(1)$ in more detail, and then use its properties to deduce perhaps the main theorem of this essay- the fact that the j-invariant of an elliptic curve with complex multiplication is an algebraic integer. In motivating the proof of this result we take a lengthy detour into some congruence subgroups of $\Gamma(1)$ and the associated modular curves. We also encounter for the first time the famous q-expansion

$$j(\tau) = \frac{1}{q} + 744 + 196884q + \dots$$

Chapter 3 is largely devoted to modular forms. I offer my own derivation of the formulæ for the dimensions of $\mathcal{M}_{2k}$ and $\mathcal{M}_{2k}^0$ (the spaces of forms and cusp forms of weight $2k$). Next Jacobi's celebrated product expansion

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

is derived, and then used to prove some bounds on the coefficients $c(n)$ in the q-expansion of j. Finally I offer a fairly non-calculational proof that

$$\left| e^{\pi\sqrt{163}} - [\![e^{\pi\sqrt{163}}]\!] \right| < 10^{-12}.$$

I have endeavoured to present the subject from an original angle, but it is inevitable that the following has a fair amount in common with existing accounts. The two books of Silverman, the account of McKean and Moll and the book by Cox are the major references.

I assume only undergraduate knowledge of Algebraic Geometry, Riemann Surfaces, Number Theory and Analysis.

Chapter 1

In this first chapter we discuss some of the theory of elliptic curves over $\mathbb{C}$. This being a large theory, we only sketch many proofs. We will look at these curves from two different points of view. These will be linked together with the help of the so-called j-invariant, whose basic properties we will discover.

1.1 Introduction to Elliptic Curves

Definition 1 *An* elliptic curve *over* $\mathbb{C}$ *is an algebraic curve over* $\mathbb{C}$ *of genus 1.*

This form being perfectly useless for most calculations, we instead use the following

Definition 2 *An elliptic curve over* $\mathbb{C}$ *is a projective curve* $E \subseteq \mathbb{P}^2(\mathbb{C})$ *given by a* Weierstraß Equation $Y^2Z = 4X^3 - g_2XZ^2 - g_3Z^3$ *where* $\Delta = g_2^3 - 27g_3^2 \neq 0$.

Throughout the following "Elliptic Curve" will mean "Elliptic Curve over $\mathbb{C}$" unless otherwise stated. Some of our results, however, are valid much more generally. The fact that Definitions 1 and 2 are essentially equivalent follows from the Riemann-Roch Theorem and a little manipulation. However we omit this and take Definition 2 as our definition. It is easy to check that a curve E given by Weierstaß equations is smooth exactly when $F(x) = 4x^3 - g_2x - g_3$ and $F'(x)$ have no common roots, i.e. precisely when the discriminant Δ is non-zero. It can also be checked that the differential $\omega = \frac{dx}{y}$ has zero valuation at all $P \in E$, including ∞. Therefore the canonical divisor class K_E is that of the zero divisor, and the genus of E really is 1.

1.1.1 The Group Law

An important property of elliptic curves is the existence of a *group law*. Using the Riemann-Roch Theorem one can show that if E is an elliptic curve with distinguished point O then the map $\Phi : E \longrightarrow Pic^0(E)$ given by

$\Phi(P) = [P - O]$ is an isomorphism. This induces an Abelian group structure on E. Furthermore it can be shown, again using Riemann-Roch, that for an elliptic curve in $\mathbb{P}^2$ this group law has an elegant geometric interpretation.

Geometric Form of Group Law. Let $Q_1, Q_2 \in \mathbb{C}$. Suppose $Q_1 Q_2$ intersects E again at R. Let OR meet E again at S. Then the group law is given by $Q_1 \oplus Q_2 = S$. Of course, we have to interpret "intersects again" slightly differently in degenerate cases (for example when $Q_1 = Q_2$ we take "the line $Q_1 Q_2$" to be "the tangent to E at Q_1")

Note that it is far from obvious, when stated geometrically, that the group law is associative. Sometimes an elliptic curve is *defined* as a genus 1 curve E together with a distinguished point O. This we denote by a pair (E, O).

1.1.2 The Endomorphism Ring

Throughout mathematics the study of structure-preserving maps between objects facilitates study of the objects themselves. A very important object associated with an elliptic curve is thus its *Endomorphism Ring*, whose properties we now outline. We will prove many of these properties later when we have an alternative characterisation of elliptic curves over $\mathbb{C}$.

Definition 3 *Let (E_1, O_1) and (E_2, O_2) be two elliptic curves. An* isogeny *is a morphism $\phi : E_1 \to E_2$ with $\phi(O_1) = O_2$. in the case $E = E_1 = E_2$, $O = O_1 = O_2$ an isogeny $\phi : E \to E$ is called an* endomorphism. *The set of such endomorphisms is denoted* $\mathrm{End}(E)$.

Now it turns out that the group law $\oplus : E \times E \to E$ is itself a morphism, and so that we may define an addition in $\mathrm{End}(E)$ by $(\phi + \psi)(P) = \phi(P) \oplus \psi(P)$. Furthermore it transpires that any isogeny of elliptic curves is also a group homomorphism. It follows that $\mathrm{End}(E)$ can be given the structure of a ring, the "multiplication" operation in this ring being composition of endomorphisms. This expains the term endomorphism ring.

$\mathrm{End}(E)$ always contains the *multiplication-by-m* maps $[m] : P \to P \oplus \ldots \oplus P$ for each $m \in \mathbb{Z}$. $\mathrm{End}(E)$ is a torsion-free $\mathbb{Z}$-module (Abelian group) with respect to addition. Indeed if $[m] \circ \phi = 0$ then $\deg([m]) \deg(\phi) = 0$ and so either ϕ or $[m]$ is constant. If ϕ is constant then it must be the zero map with constant value O; and one can check using explicit equations that $[m]$ is constant only for $m = 0$.

It may be the case that the muliplication-by-m maps form the entire endomorphism ring, so that $\mathrm{End}(E) \cong \mathbb{Z}$. If $\mathrm{End}(E)$ is larger than $\mathbb{Z}$ we say that E has *complex multiplication*. It is this situation that will interest us the most. For elliptic curves over $\mathbb{C}$, $\mathrm{End}(E)$ is always commutative with

respect to composition and has rank at most 2 as a $\mathbb{Z}$-module. Over more general fields this rank can rise to 4 and $\mathrm{End}(E)$ need not be commutative.

1.2 Elliptic Curves as Complex Tori

Now an irreducible non-singular algebraic curve C over $\mathbb{C}$ has the natural structure of a compact connected Riemann Surface. We have a Hausdorff Topology induced from the underlying projective space $\mathbb{P}^n$, and about each $P \in C$ we take as a co-ordinate chart ϕ_P the evaluation map at some local parameter t_P. The fact that t_P is a local parameter at P means that ϕ_P is a local homeomorphism; and the "transition maps" $\phi_Q \circ \phi_P^{-1} : \mathbb{C} \to \mathbb{C}$ are all analytic since the t_P are rational functions. in addition it can be shown (using two versions of the Riemann-Hurwitz formula) that the topological genus of the resulting surface is just the geometric genus of C.

We pause to be a little more explicit in the case of an elliptic curve E given by Weierstraß equation $y^2 = 4x^3 - g_2x - g_3$ in the affine plane. Co-ordinate charts are as follows:

$$\begin{array}{ll} \text{(i)} & \phi_P = x(P) \text{ for all } P = (x, y) \text{ with } y \neq 0 \\ \text{(ii)} & \phi_P = y(P) \text{ for all } P = (x, y) \text{ with } 12x^2 - g_2 \neq 0 \\ \text{(iii)} & \phi_P = \frac{x}{y}(P) \text{ at } \infty \end{array} \tag{1.1}$$

Any rational function $f \in \mathbb{C}(E)$ is then easily seen to provide a meromorphic function on the Riemann Surface E. In fact the converse is also ture, as we'll see later.

So an elliptic curve is simply a complex torus. Now we know another way in which complex tori arise- they are quotients of $\mathbb{C}$ by a lattice Λ (or more correctly, by the action on $\mathbb{C}$ of a free group on independant generators ω_1, ω_2). Moreover in this context we have a lot of information- we will show how to construct all meromorphic functions on $\mathbb{C}/\Lambda$, for example. The most important such function is the Weierstraß function $\wp$ defined by

$$\wp(z; \Lambda) = \frac{1}{z^2} + \sum_{\omega \in \Lambda^*} \left\{ \frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right\} \tag{1.2}$$

This is a meromorphic function of valency two, i.e. $\wp : \mathbb{C}/\Lambda \to \mathbb{C}_\infty$ is a double cover. More importantly it turns out to be relatively easy to describe the endomorhism ring of an elliptic curve in the context of lattices. It would therefore be nice if all elliptic curves arose from this construction. This truns out to be the case, a reult known as the Uniformization Theorem for Elliptic Curves. It also turns out that every torus $\mathbb{C}/\Lambda$ is an elliptic curve in a natural way. We study this first, as it paves the way for the converse result.

Recall first that we have

$$\wp'(z)^2 - 4\wp(z)^3 - g_2(\Lambda)\wp(z) - g_3(\Lambda) \tag{1.3}$$

where $g_2(\Lambda)$, $g_3(\Lambda)$ are defined by

$$g_2(\Lambda) = 60 \sum_{\omega \in \Lambda^*} \omega^{-4},$$

$$g_3(\Lambda) = 140 \sum_{\omega \in \Lambda^*} \omega^{-6}.$$

This can be checked using the following easily derived Laurent Expansion of $\wp_\Lambda$ about 0.

Proposition 4 (Laurent Expansion of $\wp_\Lambda$ about 0) *For $r \geq 3$ define $G_r(\Lambda) = \sum_{\omega \in \Lambda^*} \omega^{-r}$ (so that $g_2 = 60G_4$, $g_3 = 140G_6$). then we have the Laurent Expansion*

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

about 0.

Proof For $|x| < 1$ one has $(1-x)^{-2} = 1 + \sum_{n=1}^{\infty}(n+1)x^n$. Hence if $|z| < |\omega|$ we can put $x = \frac{z}{\omega}$ to get $(z-\omega)^{-2} - \omega^{-2} = \sum_{n=1}^{\infty} \frac{n+1}{\omega^{n+2}} z^n$. Summing over Λ^* and interchanging the order of summation, we get the result. This *is* valid because the necessary sums converge absolutely; see the proof of Theorem 16 for an indication of why this is true.

To check (1.3) one shows that $\wp'(z)^2 - 4\wp(z)^3 - g_2\wp(z) - g_3$ is analytic with value 0 at $z = 0$ by using the Laurent Expansion we have just obtained. Since then it is doubly periodic and analytic on $\mathbb{C}$, it follows from a standard argument involving Liouville's Theorem that it must be constant.

It appears from (1.3) that we have a map ϕ from $\mathbb{C}/\Lambda$ to an elliptic curve E_Λ with affine Weierstraß equation $y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$. ϕ is given by

$$\phi(z) = (\wp(z), \wp'(z)).$$

This map has some remarkable properties.

Theorem 5 (i) *E_Λ really is an elliptic curve, i.e. $\Delta(\Lambda) = g_2^3(\Lambda) - 27g_3^2(\Lambda) \neq 0$.*

(ii) *ϕ is an isomorphism of Riemann Surfaces*

(iii) *The meromorphic functions on $\mathbb{C}/\Lambda$ correspond under ϕ to the rational*

functions on E_Λ

(iv) *The group law on* (E_Λ, ∞) *lifts under* ϕ *to ordinary vector addition in the quotient* $\mathbb{C}/\Lambda$.

Proof (i) We have already remarked that $\Delta(\Lambda)$ is the discriminant of the polynomial $F(x) = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$, and so it suffices to show that the roots e_1,e_2,e_3 of F are all distinct. Now

$$\wp'(z) = 4(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3). \tag{1.4}$$

But $\wp'$ is an odd Λ-elliptic function and so either vanishes or has a pole at all points of $\frac{1}{2}\Lambda$. Hence

$$\wp'(\frac{1}{2}\omega_1) = \wp'(\frac{1}{2}\omega_2) = \wp'(\frac{1}{2}(\omega_1 + \omega_2)) = 0.$$

At these points, then, $\wp$ has multiplicity at least two. But $\wp : \mathbb{C}/\Lambda \to \mathbb{C}_\infty$ is a double cover, and so $\wp$ has multiplicity exactly two and the values $\wp(\frac{1}{2}\omega_1)$, $\wp(\frac{1}{2}\omega_2)$ and $\wp(\frac{1}{2}(\omega_1 + \omega_2))$ are all distinct. By (1.4), these must be e_1, e_2 and e_3 in some order; in particular e_1, e_2 and e_3 are all distinct.

(ii) Note that $\wp(z)$ and $\wp'(z)$ are both analytic except for poles at the points of Λ. It follows that ϕ is analytic with respect to the obvious chart on $\mathbb{C}/\Lambda$ and the charts (1.1) on E_Λ, except possibly for the point ∞. However ϕ is also analytic here since $\wp(z)/\wp'(z)$ is holomorphic at $z = 0$. To prove that ϕ is an isomorphism we use the fact that an analytic map between compact Riemann Surfaces has a well-defined degree, and that this degree is 1 precisely if the map is an isomorphism. Indeed an analytic map is already locally analytically invertible and the degree 1 condition ensures global invertibility. Now we have a commutative triangle

$$\begin{array}{ccc} \mathbb{C}/\Lambda & \xrightarrow{\psi} & E_\Lambda \\ & \wp \searrow \quad \swarrow \pi_x & \\ & \mathbb{C}_\infty & \end{array}$$

where π_x is projection onto the x co-ordinate. Both $\wp$ and π_x are analytic maps of degree 2 and so $\deg(\phi) = \deg(\wp)/\deg(\pi_x) = 1$.

(iii) Now that we have our isomorphism ϕ, it suffices to show that the only meromorphic functions on the complex torus $\mathbb{C}/\Lambda$ are rational functions in $\wp(z)$ and $\wp'(z)$. Regard a meromorphic function on $\mathbb{C}/\Lambda$ as a doubly periodic function on $\mathbb{C}$. We prove the result first in the case f even. Choose $c, d \in \mathbb{C}$ such that the roots z of $f(z) = c$ and $f(z) = d$ are all simple and are such that $2z \notin \Lambda$. Let these two sets of roots be $\{a_1, -a_1, a_2, -a_2, \ldots a_n, -a_n\}$ and $\{b_1, -b_1, \ldots b_n, -b_n\}$. then the function

$$g(z) = \frac{f(z) - c}{f(z) - d}$$

has simple zeros at precisely $\pm a_i$ and simple poles just at $\pm b_j$. But

$$h(z) = \frac{\prod_i (\wp(z) - \wp(a_i))}{\prod_i (\wp(z) - \wp(b_j))}$$

also has exactly these propeties, and hence g/h is a doubly-periodic meromorphic function with no poles. By the standard application of Liouville's Theorem, it must be constant. It follows immediately that $f \in \mathbb{C}(\wp)$. If we drop the assumption that f is even then we can write $f = f_e + f_o$ where $f_e = \frac{1}{2}\{f(z) + f(-z)\}$ and $f_o = \frac{1}{2}\{f(z) - f(-z)\}$. Now $\wp'$ is an odd function and so $\wp' \cdot f_o$ is even. Since both f_e and $\wp' \cdot f_o$ are also both doubly-periodic, the result follows immediately.

(iv) There are at least two ways to prove this. One is to ask what identity $\wp$ would have to satisfy for the result to be true, and then prove it. This we do by working out explicit equations for the group law on E_Λ (using the geometric form) and pulling back to $\mathbb{C}/\Lambda$ using ϕ. The answer is that $\wp$ must satisfy

$$\wp(z+w) = -\wp(z) - \wp(w) + \frac{1}{4}\left(\frac{\wp'(z) - \wp'(w)}{\wp(z) - \wp(w)}\right)^2 \tag{1.5}$$

whenever $z, w, z+w \notin \Lambda$. This *can* be proved by a careful consideration of Laurent expansions. We, however, will give a more elegant proof. (1.5) can then be recovered, if desired, simply by applying ϕ. We require two Lemmas.

Lemma 6 *Suppose $n_1, \ldots n_r \in \mathbb{Z}$ and $z_1, \ldots z_r \in \mathbb{C}$ satisfy $\sum n_i = 0$ and $\sum n_i z_i \in \Lambda$. Then there is an ellpitic function f on Λ (that is, a Λ-doubly periodic function) whose divisor (f) of poles and zeros is precisely $\sum n_i[z_i]$.*

Proof We may assume $\sum n_i z_i = 0$. The required function f is given by the formula

$$f(z) = \prod_i \sigma(z - z_i)^{n_i}.$$

Here

$$\sigma(z;\Lambda) = z \prod_{\omega \in \Lambda^*} \left(1 - \frac{z}{\omega}\right) e^{-(z/\omega) - \frac{1}{2}(z/\omega)^2}$$

is the Weierstaß σ-function and has two key properties for us:

Property 1: σ has simple zeros at the points of Λ and nowhere else

Property 2: Let $\omega \in \Lambda$. then $\sigma(z+\omega) = e^{az+b}\sigma(z)$ for some $a = a(\omega)$, $b = b(\omega)$ and all $z \in \mathbb{C}$.

Property 1 shows that $(f) = \sum n_i[z_i]$ as desired. Property 2 gives

$$\begin{aligned} f(z+\omega) &= \prod_i e^{n_i(a(z-z_i)+b)} \cdot f(z) \\ &= f(z) \end{aligned}$$

It is of some interest that the conditions of this Lemma are also necessary, and so we have again constructed the most general elliptic function with respect to the lattice Λ.

Lemma 7 *Let (E, O) be an elliptic curve, and let $D = \sum n_P P$ be a divisor on E. Then if D is principal we have $\bigoplus n_P P = 0$, where $\oplus$ denotes addition in the group law on E.*

Proof Recall that the group law can be defined by the map Ψ sending P to $[P - O]$. If $(f) = \sum n_P P$ then $\sum n_P = 0$ and so

$$\Psi\left(\sum n_P P\right) = \left[\sum n_P P\right] = [(f)] = [O] = \Psi(0),$$

i.e. $\bigoplus n_P P = O.//$

To deduce Theorem 5 (iv), let $z_1, z_2 \in \mathbb{C}$. Choose an elliptic function f with $(f) = [z_1 + z_2] - [z_1] - [z_2] + [0]$. That this is possible follows from Lemma 6. We know from (iii) that $f = \phi^* R$ for some rational function R. But $\phi : \mathbb{C}/\Lambda \to E_\Lambda$ is an isomorphism of Riemann Surfaces and so

$$v_{\phi(P)}(R) = v_P(\phi^* R) = v_P(f)$$

for any $P \in \mathbb{C}/\Lambda$. It follows that

$$(R) = [\phi(z_1 + z_2)] - [\phi(z_1)] - [\phi(z_2)] + [\phi(0)]$$

and so, by Lemma 7, that $\phi(z_1 + z_2) = \phi(z_1) \oplus \phi(z_2)$ relative to the group law on (E_Λ, ∞). This concludes the proof of Theorem 5.//

We take the opportunity to note now that the choice of regular differential $\omega = \frac{dx}{y}$ on E_Λ now seems a little less arbitrary; in fact

$$\phi^* \omega = \frac{d\wp(z)}{\wp'(z)} = dz$$

is the most natural differential of all on $\mathbb{C}/\Lambda$.

1.3 Structure of $\mathrm{End}(E)$

The next question of interest is to ask what the endomorphism ring $\mathrm{End}(E_\Lambda)$ looks like in terms of the Riemann Surface structure. the simple answer is given by the following Theorem, which may be considered the main justification for wishing to regard Elliptic Curves as quotients of $\mathbb{C}$ by lattices.

Theorem 8 (i) *Let (E_1, O) and (E_2, O) be two elliptic curves corresponding to lattices $\mathbb{C}/\Lambda_1$ and $\mathbb{C}/\Lambda_2$. Let ϕ_1, ϕ_2 be the corresponding isomorphisms (cf Theorem 5). Then in the commutative square*

$$\begin{array}{ccc} \mathbb{C}/\Lambda_1 & \xrightarrow{g} & \mathbb{C}/\Lambda_2 \\ \downarrow{\scriptstyle \psi_1} & & \downarrow{\scriptstyle \psi_2} \\ E_1 & \xrightarrow{f} & E_2 \end{array}$$

g is an analytic map with $g(0) = 0$ if and only if f is an isogeny. In essence, isogenies and analytic maps fixing 0 are the same thing.

(ii) *The analytic maps $g : \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2$ with $g(0) = 0$ are precisely those induced from multiplication maps $m_\alpha : z \to \alpha z$ on $\mathbb{C}$, where $\alpha\Lambda_1 \subseteq \Lambda_2$.*

Proof (i) An isogeny is just a rational map. Hence it is fairly clear from the form of the charts (1.1) that if f is an isogeny then g is analytic. The converse, however, requires a little more work. Any analytic map $g : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda_2$ has a "lift" $\tilde{g} : \mathbb{C} \to \mathbb{C}$. For our purposes this is an analytic map with $\tilde{g}(0) = 0$ which makes the following diagram commute-

$$\begin{array}{ccc} \mathbb{C} & \xrightarrow{\tilde{g}} & \mathbb{C} \\ \downarrow{\scriptstyle \pi_1} & & \downarrow{\scriptstyle \pi_2} \\ \mathbb{C}/\Lambda_1 & \xrightarrow{g} & \mathbb{C}/\Lambda_2 \end{array}$$

The proof that such a map exists uses the simple-connectedness of $\mathbb{C}$ and the so-called "monodromy theorem". See Jones and Singerman [7] for details. Now for any fixed $\omega \in \Lambda_1$ we have

$$\pi_2(\tilde{g}(z + \omega)) = g(\pi_1(z + \omega)) = g(\pi_1(z)) = \pi_2(\tilde{g}(z)). \tag{1.6}$$

The map f satisfies

$$f\left(\wp(z;\Lambda_1), \wp'(z;\Lambda_1)\right) \;=\; \left(\wp(\tilde{g}(z);\Lambda_2), \wp'(\tilde{g}(z);\Lambda_2)\right).$$

But (1.6) guarantees that $\wp(\tilde{g}(z);\Lambda_2)$ and $\wp'(\tilde{g}(z);\Lambda_2)$ are Λ_1-elliptic, and so by Theorem 5 (ii) they are rational functions of $\wp(z;\Lambda_1)$ and $\wp'(z;\Lambda_1)$.

(ii) $\tilde{g}$ is continuous and Λ_2 is discrete, so from (1.6) we have $\tilde{g}(z+\omega) - \tilde{g}(z) = c_\omega$ for some constant c_ω. Hence $\tilde{g}'$ is Λ_1-elliptic. But it is also analytic, and so by the usual argument involving Liouville's theorem it is constant. Since $\tilde{g}(0) = 0$ we must have $\tilde{g} = m_\alpha$ for some α. Applying (1.6) once again shows that $\alpha\Lambda_1 \subseteq \Lambda_2$. Conversely it is easy to see that any such map does give an analytic map from $\mathbb{C}/\Lambda_1$ to $\mathbb{C}/\Lambda_2$.//

Apart from proving the Uniformization Theorem, we have now fully set up the correspondence between Elliptic Curves over $\mathbb{C}$ and quotients $\mathbb{C}/\Lambda$. Let us, however, state the Uniformization theorem and look at some consequences concerning endomorphism rings in the light of Theorems 5 and 8.

Theorem 9 (Uniformisation Theorem for Elliptic Curves over $\mathbb{C}$) *Every isomorphism class of elliptic curves over $\mathbb{C}$ arises as some quotient $\mathbb{C}/\Lambda$.*

Proof See later for the full details. What we are required to show is that if γ_2, γ_3 satisfy $\gamma_2^3 - 27\gamma_3^2 \neq 0$ then we can find a lattice Λ with $\gamma_2 = g_2(\Lambda)$, $\gamma_3 = g_3(\Lambda)$. The curve with Weierstraß Equation $Y^2Z = X^3 - \gamma_2 XZ^2 - \gamma_3 Z^3$ will then be isomorphic to $\mathbb{C}/\Lambda$.//

From now on we supress the explicit isomorphism ϕ of Theorem 5 and regard $\mathbb{C}/\Lambda$ and E_Λ as "the same".

1.3.1 Consequences of Theorem 9

Consequence 9.1 (E_1, O_1) and (E_2, O_2) are isogenous (i.e. there exists a non-trivial isogeny $\phi : E_1 \to E_2$) if and only if there is $\alpha \in \mathbb{C}^*$ with $\alpha\Lambda_1 \subseteq \Lambda_2$. (E_1, O_1) and (E_2, O_2) are isomorphic exactly when Λ_1 and Λ_2 are homothetic, i.e. when $\alpha\Lambda_1 = \Lambda_2$ for some $\alpha \in \mathbb{C}$.

Consequence 9.2 Description of $\mathrm{End}(E_\Lambda)$. From Theorem 8,

$$\mathrm{End}(E_\Lambda) \cong \{\alpha \in \mathbb{C} | \alpha\Lambda \subseteq \Lambda\}.$$

In particular, $\mathrm{End}(E_\Lambda)$ is a commutative torsion-free $\mathbb{Z}$-module. The exact nature of $\mathrm{End}(E_\Lambda)$ depends on the lattice Λ, although only the homothety class of Λ is important by Consequence 9.1 above. Since any lattice is homothetic to one of the form $\Lambda_\tau = \mathbb{Z} \oplus \mathbb{Z}\tau$ ($\tau \notin \mathbb{R}$) it suffices to consider this case. Which α are allowable? Since $\alpha \cdot 1 \in \Lambda$ one has $\alpha = a + b\tau$ for some $a, b \in \mathbb{Z}$. It is then necessary and sufficient that $\alpha \cdot \tau \in \Lambda$, i.e. that

$$(a + b\tau)\tau = c\tau + d \tag{1.7}$$

for some $c, d \in \mathbb{Z}$. Unless τ is a quadratic irrational, then, one has $b = 0$ and so $\alpha \in \mathbb{Z}$. In this case $\mathrm{End}(E_\Lambda) \cong \mathbb{Z}$: this accounts for "almost all" elliptic curves.

If τ *is* a (purely imaginary) quadratic irrational satisfying a minimal equation $A\tau^2 + B\tau + C = 0$ over $\mathbb{Z}$, where we may assume $\gcd(A, B, C) = 1$, then it is not hard to see from 1.7 that $\alpha\Lambda \subseteq \Lambda \Leftrightarrow A|b$. Indeed we require that there exist e, f with $b\tau^2 + e\tau + f = 0$, and so $A|bB$ and $A|bC$. However

there are u, v, w such that $Au + Bv + Cw = 1$, and so in fact $A|b$. Hence in this case $\mathrm{End}(E_\Lambda) \cong \mathbb{Z} \oplus \mathbb{Z}A\tau$. Now $A\tau$ is an algebraic in $\mathbb{Q}(\tau)$ and so $\mathcal{O} \cong \mathrm{End}(E_\Lambda)$ satisfies the conditions required to be an *order* in $\mathbb{Q}(\tau)$.

By saying that $\mathcal{O}$ is an order in a number field K we mean

(i) $\mathcal{O}$ is a subring of K containing 1

(ii)$\mathcal{O}$ is a finitely-generated $\mathbb{Z}$-module

(iii) $\mathcal{O} \otimes_{\mathbb{Q}} \mathbb{Q} = K$.

Conditions (i) and (ii) imply that $\mathcal{O} \subseteq \mathcal{O}_K$, the ring of integers of K. To summarise,

Theorem 10 *Let E_Λ be an elliptic curve. Let $\iota : \mathrm{End}(E_\Lambda) \to \mathbb{C}$ be the map described in Theorem 8, i.e. $\iota(\mathrm{End}(E_\Lambda)) = \{\alpha \in \mathbb{C} | \alpha\Lambda \subseteq \Lambda\}$. Then either*

(i) $\iota(\mathrm{End}(E_\Lambda)) = \mathbb{Z}$ *or*
(ii) $\iota(\mathrm{End}(E_\Lambda))$ *is an order $\mathcal{O}$ in a quadratic field K over $\mathbb{Q}$.*

In the latter case E_Λ is said to admit "complex multiplication".

Consequence 9.3 It is interesting to ask the reverse question. Namely, given an order $\mathcal{O}$ in a quadratic field K, when is $\mathrm{End}(E_\Lambda) \cong \mathcal{O}$? For simplicity (to avoid setting out the theory of orders) we treat only the case $\mathcal{O} = \mathcal{O}_K$. We may suppose that $K = \mathbb{Q}(\tau)$ where $\{1, \tau\}$ is an integral basis for $\mathcal{O}_K$. Now given an abstract isomorphism $\mathrm{End}(E_\Lambda) \cong \mathcal{O}_K$, all we know is that $\iota(\mathrm{End}(E_\Lambda))$ is a subring of $\mathbb{C}$ abstractly isomorphic to $\mathcal{O}_K$. Our first lemma removes this subtlety.

Lemma 11 *If $R \subseteq \mathbb{C}$ is abstractly isomorphic to $\mathcal{O}_K$, then $R = \mathcal{O}_K$ as sets.*

Proof Let $\phi : \mathcal{O}_K \to R$ be the isomorphism. By the definition of a ring isomorphism, ϕ fixes $\mathbb{Z}$. Now any $\eta \in \mathcal{O}_K \setminus \mathbb{Z}$ satisfies an irreducible monic quadratic $\eta^2 + p\eta + q = 0$ over $\mathbb{Z}$. Furthermore the other root $\eta' = -p - \eta$ also lies in $\mathcal{O}_K$. Applying ϕ to this quadratic shows that $\{\phi(\eta), \phi(\eta')\} = \{\eta, \eta'\}$. So ϕ (at worst) swaps conjugate quadratic integers in $\mathcal{O}_K$, and $R = \mathcal{O}_K$ as sets.//

The next important thing to note is that we may restrict attention to sublattices of $\mathcal{O}_K = \mathbb{Z} \oplus \mathbb{Z}\tau$.

Lemma 12 *If $\mathrm{End}(E_\Lambda) \cong \mathcal{O}_K$, then Λ is homothetic to a sublattice of $\mathcal{O}_K$.*

Proof It follows from Lemma 11 that $\iota(\mathrm{End}(E_\Lambda)) = \mathcal{O}_K$. Suppose $\Lambda = \mathbb{Z} \oplus \mathbb{Z}\eta$. Then $\tau \cdot 1 \in \Lambda$, and so $\eta = \frac{1}{b}(\tau - a)$ for some $a, b \in \mathbb{Z}$. But then Λ

is homothetic to $\mathbb{Z}b \oplus \mathbb{Z}(\tau - a)$.//

Now a sublattice of $\mathcal{O}_K$ which is closed under multiplication by $\mathcal{O}_K$ is none other than an ideal of $\mathcal{O}_K$. In fact we have the following result.

Lemma 13 $\mathrm{End}(E_\Lambda) \cong \mathcal{O}_K$ *if and only if* Λ *is homothetic to an ideal* $\Lambda' \triangleleft \mathcal{O}_K$.

Proof We have done all of this except for showing that if $\Lambda' \triangleleft \mathcal{O}_K$ then $\mathrm{End}(E_\Lambda)$ is no bigger than $\mathcal{O}_K$. Suppose $\Lambda' = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$ and that $\alpha \in \iota(\mathrm{End}(E_\Lambda))$. Then $\alpha\omega_1 = c\omega_1 + d\omega_2$ for some $c, d \in \mathbb{Z}$ and so we certainly have $\alpha \in \mathbb{Q}(\omega_1, \omega_2) \subseteq \mathbb{Q}(\tau)$. But Theorem 10 shows that α is an algebraic integer, and so $\alpha \in \mathcal{O}_K$.//

When do two ideals $\Lambda, \Lambda' \triangleleft \mathcal{O}_K$ give rise to isomorphic curves $E_\Lambda, E_{\Lambda'}$? The answer is, almost by definition, that $E_\Lambda \cong E_{\Lambda'}$ precisely when Λ and Λ' are in the same ideal class in the ideal class group $\mathcal{C}(\mathcal{O}_K)$. Indeed

Theorem 14 *We have a bijective correspondence between the isomporhism classes of elliptic curves over* $\mathbb{C}$ *with* $\mathrm{End}(E) \cong \mathcal{O}_K$ *and* $\mathcal{C}(\mathcal{O}_K)$. *In particular, the number of isomorphism classes is precisely the class number* h_K.

Consequence 9.4 Degrees of Maps. If we think geometrically, it is clear that the degree of an isogeny corresponding to the multiplication map m_α is $|\alpha|^2$. Hence the degree of the multiplication-by-m map $[m]$ is m^2. Furthermore (from Consequence 9.2) suppose $\Lambda = \mathbb{Z} \oplus \mathbb{Z}\tau$ with $A\tau^2 + B\tau + C = 0$ and $\gcd(A, B, C) = 1$. Then the possible degrees of maps $\phi : E_\Lambda \to E_\Lambda$ are the values taken by the monic binary quadratic form $f(x, y) = x^2 + xy(\nu + \overline{\nu}) + y^2\nu\overline{\nu}$ where $\nu = A\tau$. in particular the set of values taken by f as x, y range over $\mathbb{Z}$ is multiplicatively closed and we could easily derive the form of r, s such that $f(x_1, y_1)f(x_2, y_2) = f(r, s)$.

1.4 The j-invariant

To prove Theorem 9 we need to introduce an object which will occupy the rest of the essay- the so-called j-invariant. This will be a complex number assigned to each elliptic curve, with the property that

$$j(E) = j(E') \Leftrightarrow E \cong E' \tag{1.8}$$

There are of course uncountably many functions with this property, but our j will have remarkable analytic and arithmetic properties which will become gradually more apparent as we proceed. Let us start by defining j for a curve of form $\mathbb{C}/\Lambda$.

1.4.1 j-invariant of a lattice

At the moment we have three quantities associated to a lattice, namely

$$g_2(\Lambda) = 60 \sum_{\omega \in \Lambda^*} \omega^{-4},$$

$$g_3(\Lambda) = 140 \sum_{\omega \in \Lambda^*} \omega^{-6}$$

and

$$\Delta(\Lambda) = g_2^3(\Lambda) - 27 g_3^2(\Lambda).$$

None of these are homothety invariants; indeed one has for any $\lambda \in \mathbb{C}^*$ $g_2(\lambda\Lambda) = \lambda^{-4} g_2(\Lambda)$, $g_3(\lambda\Lambda) = \lambda^{-6} g_3(\Lambda)$ and $\Delta(\lambda\Lambda) = \lambda^{-12}\Delta(\Lambda)$. This enables us to construct a great many homothety invariants. However we have the additional information (Theorem 5(i)) that for any lattice Λ, $\Delta(\Lambda) \neq 0$. This suggests that we might consider one of the invariants g_2^3/Δ or g_3^2/Δ. In fact we define

$$j(\Lambda) = 1728 \frac{g_2(\Lambda)^3}{\Delta(\Lambda)}.$$

The factor of 1728 appears for reasons which will become apparent in due course.

By a slight abuse of notation we also define

$$j(g_2, g_3) = \frac{1728 g_2^3}{g_2^3 - 27 g_3^2}$$

for any pair $(g_2, g_3) \in \mathbb{C} \times \mathbb{C}$ with $g_2^3 - 27 g_3^2 \neq 0$. We would like to further define, for an elliptic curve $E(g_2, g_3)$ with affine Weierstraß equation $y^2 = 4x^3 - g_2 x - g_3$, $j(E)$ to be $j(g_2, g_3)$. This would certainly fit with our existing definition in the case $E \cong \mathbb{C}/\Lambda$. However with our current state of knowledge this is not well-defined, as we do not know enough about when $E(g_2, g_3) \cong E(g_2', g_3')$. Nevertheless we do have the following result.

Lemma 15 *Suppose that $j(g_2, g_3) = j(g_2', g_3')$ (where $g_2^3 - 27 g_3^2$, $g_2'^3 - 27 g_3'^2 \neq 0$). Then $E(g_2, g_3) \cong E(g_2', g_3')$.*

Proof If $g_2', g_3' \neq 0$ we have $\left(\frac{g_3}{g_3'}\right)^2 = \left(\frac{g_2}{g_2'}\right)^3$ and so we can write $g_3' = g_3 c^3$, $g_2' = g_2 c^2$ for some $c \in \mathbb{C}$. But then we have an isomorphicm $\phi : E(g_2, g_3) \to E(g_2', g_3')$ given by $(x, y) \xrightarrow{\phi} (cx, \pm c^{3/2} y)$. If $g_2' g_3' = 0$ the analysis is similar.//

It is easy to see from this that to prove the Uniformization Theorem (Theorem 9) it suffices to show that j : lattices $\to \mathbb{C}$ is surjective. Indeed given

$E(g_2, g_3)$ we can then choose Λ with $j(g_2, g_3) = j(g_2(\Lambda), g_3(\Lambda))$ and Lemma 15 guarantees

$$E(g_2, g_3) \cong E(g_2(\Lambda), g_3(\Lambda)) \cong \mathbb{C}/\Lambda.$$

Surjectivity of j is therefore our objective.

1.4.2 Surjectivity of j

In order to prove that j : lattices $\to \mathbb{C}$ is surjective we look at j in a slightly different light. Recall that $j(\Lambda)$ depends only on the homothety type of Λ, and that every lattice is homothetic to one of form $\Lambda_\tau = \mathbb{Z} \oplus \mathbb{Z}\tau$ for some $\tau \in \mathbb{H}$, the open upper half-plane. Hence we can, by yet another abuse of notation, define a function $j : \mathbb{H} \to \mathbb{C}$ by setting $j(\tau) = j(\Lambda_\tau)$. In this context we have the weapon of complex analysis at our disposal. Indeed one has the following important result.

Theorem 16 $j : \mathbb{H} \to \mathbb{C}$ *is analytic.*

Proof Since $j(\tau) = 1728 g_2^3(\tau)/\Delta(\tau)$ (with an obvious notation) and $\Delta(\tau) \neq 0$, it suffices to show that $g_2(\tau)$ and $g_3(\tau)$ are analytic. Recalling that

$$g_2(\tau) = 60 \sum_{\substack{m,n \in \mathbb{Z} \\ (m,n) \neq (0,0)}} (m + n\tau)^{-4},$$

$$g_3(\tau) = 140 \sum_{\substack{m,n \in \mathbb{Z} \\ (m,n) \neq (0,0)}} (m + n\tau)^{-6}$$

we proceed much as in the standard proof that $\wp(z)$ is meromorphic. Indeed the sums for g_2 and g_3 converge absolutely and uniformly on compacta. To see this let $D \subseteq \mathbb{H}$ be compact and consider the function $f : D \times S^1 \to \mathbb{R}$ given by $f(\tau, e^{it}) = |\cos t + \tau \sin t|$. This is continuous on the compact set $D \times S^1$ and is never zero. Hence it is bounded below by some c_D, and one has $|m + n\tau| \geq c_d \sqrt{m^2 + n^2}$ for all $(m, n) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$. But

$$\sum_{\substack{m,n \in \mathbb{Z} \\ (m,n) \neq (0,0)}} (m^2 + n^2)^{-\gamma}$$

converges whenever $\gamma > 1$.//

j is actually a very special type of analytic function, because of the fact that if Λ_τ and $\Lambda_{\tau'}$ are homothetic then $j(\tau) = j(\tau')$. The next proposition analyses exactly when this happens.

Proposition 17 (i) *Let*

$$SL_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; ad - bc = 1, a, b, c, d \in \mathbb{Z} \right\}$$

Then $SL_2(\mathbb{Z})$ acts on $\mathbb{H}$ by

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

Furthermore $\pm I$ acts trivially on $\mathbb{H}$, and so we can factor through to give an action of the modular group $\Gamma(1) = SL_2(\mathbb{Z})$ *on $\mathbb{H}$. Two lattices Λ_τ and $\Lambda_{\tau'}$ are homothetic if and only if τ and τ' lie in the same orbit under this action of $\Gamma(1)$.*

(ii) *Every $\Gamma(1)$-orbit contains a point of the region $\mathcal{F}$ defined by*

$$\mathcal{F} = \left\{ \tau \in \mathbb{H} \;\middle|\; |\tau| \geq 1 \text{ and } |\Re(\tau)| \leq \frac{1}{2} \right\}$$

(iii) *No two points in the interior $\mathcal{F}^\circ$ are in the same $\Gamma(1)$-orbit.*

Proof In the following we abuse notation by using matrices to denote elements of $\Gamma(1)$. These should always be read modulo $\pm I$.

(i) This is just a straightforward computation.

(ii) To emphasise the links between different area, we prove this and (iii) using quadratic forms. The arguments may be familiar from the reduction theory of positive-definite binary quadratic over $\mathbb{Z}$. Here, however, we work over $\mathbb{R}$. The link to quadratic forms comes from the map

$$\theta : \{\text{positive definite forms } (a, b, c) \text{ over } \mathbb{R}\} \to \mathbb{H}$$

defined by

$$\theta(a, b, c) = \text{root of } at^2 + bt + c = 0 \text{ in } \mathbb{H}$$

Importantly this map is surjective, since $\theta(1, \tau + \overline{\tau}, |\tau|^2) = \tau$ for any $\tau \in \mathbb{H}$.

Now we have a notion of equivalence of forms- this is defined in terms of unimodular transformations

$$\begin{aligned} x' &= px + qy, \\ y' &= rx + sy \end{aligned}$$

where

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} \in SL_2(\mathbb{Z}).$$

Thus if $(a', b', c') = \begin{pmatrix} p & q \\ r & s \end{pmatrix} (a, b, c)$ then

$$\begin{aligned} a' &= ap^2 + bpr + cr^2 \\ b' &= 2apq + b(qr + ps) + 2crs \\ c' &= aq^2 + bqs + cs^2. \end{aligned} \tag{1.9}$$

This action of $SL_2(\mathbb{Z})$ on forms is related to the action of $SL_2(\mathbb{Z})$ on $\mathbb{H}$ via θ in a particularly nice way. Indeed if $(a',b',c') = \begin{pmatrix} p & q \\ r & s \end{pmatrix} (a,b,c)$ then we have

$$a't^2 + b't + c = (rt+s)^2(at'^2 + bt' + c)$$

where $t' = \frac{pt+q}{rt+s}$. Hence $\theta(a',b',c') = \begin{pmatrix} p & q \\ r & s \end{pmatrix}^{-1} \theta(a,b,c)$ under the action of $SL_2(\mathbb{Z})$ on $\mathbb{H}$. In words, $SL_2(\mathbb{Z})$-equivalent forms have $SL_2(\mathbb{Z})$-equivalent roots.

Now it is easy to see which forms (a,b,c) have $\theta(a,b,c) \in \mathcal{F}$. They are what I call *pseudoreduced* forms, in which $-a \le b \le a \le c$. To prove (ii), then, it suffices to show that every form is equivalent to a pseudoreduced one. Let us single out two special types of equivalence transformation, namely

Type I: $(a,b,c) \to (c,-b,a)$

Type II: $(a,b,c) \to (a, b \pm 2a, a \pm b + c)$.

These correspond to matrices $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix}$ respectively. Starting with a form (a,b,c), use Type II transformations to get $|b'| \le |a'|$. If then $c' \ge a'$ we are done; if not, apply a Type I transformation and repeat. Clearly if this process terminates we get a pseudoreduced form equivalent to our original one, and we will have our result. To see that this is the case, note that the coefficient of x^2 is invariant under transformations of Type II and is reduced by any Type I transformation that we may be forced to use. However from (1.9) we see that any such leading coefficient lies in the set

$$\Sigma = \left\{ ap^2 + bpr + cr^2 \;\middle|\; p,r \in \mathbb{Z}, \gcd(p,r) = 1 \right\} \tag{1.10}$$

However it becomes obvious that Σ is well-ordered if we write

$$ap^2 + bpr + cr^2 = a\left(p + \frac{br}{2a}\right)^2 - \frac{\Delta}{4a} r^2,$$

where $\Delta = b^2 - 4ac < 0$. This completes the proof of (ii).

(iii) Forms (a,b,c) with $\theta(a,b,c) \in \mathcal{F}^\circ$ satisfy $-a < b < a < c$ and I call them *strictly reduced*. Now if two forms (a,b,c) and (a',b',c') are equivalent then the corresponding sets Σ and Σ' defined in (1.10) are clearly equal. So let us look at Σ in more detail, supposing that (a,b,c) and (a',b',c') are strictly reduced. We have

$$\begin{aligned} ap^2 + bpr + cr^2 &\ge a|p|^2 - |b||p||r| + c|r|^2 \\ &= a|p|^2 + c|r|(|r| - |p|) + (c - |b|)|r||p| \\ &\ge (a - |b| + c)|p|^2 \text{ if } |r| \ge |p|. \end{aligned}$$

Similarly

$$ap^2 + bpr + cr^2 \geq (a - |b| + c)|r|^2$$

if $|p| \geq |r|$, and so

$$ap^2 + bpr + cr^2 \geq (a - |b| + c)\min(p^2, r^2).$$

It follows immediately from this and the fact that (a, b, c) is strictly reduced that the three smallest elements of Σ are $a < c < a - |b| + c$. Therefore a, $|b|$ and c can be recovered uniquely from Σ, and hence $(a', b', c') = (a, \pm b, c)$. However if $(a, \pm b, c) = \begin{pmatrix} p & q \\ r & s \end{pmatrix} (a, b, c)$ then one has

$$\begin{aligned} ap^2 + bpr + cr^2 &= a \text{ and} \\ aq^2 + bqs + cs^2 &= c \end{aligned}$$

and so by comments above $\begin{pmatrix} p & q \\ r & s \end{pmatrix} = \pm I$. It follows that any two distinct strictly reduced forms (a, b, c) and (a', b', c') are inequivalent under the action of $\Gamma(1)$, and furthermore that each such form has trivial stabilizer. This concludes the proof of Proposition 17.//

Let us note that out of our proof we get the following result about $\Gamma(1)$.

Corollary 18 *$\Gamma(1)$ is generated by matrices $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$.*

Proof It follows from Proposition 17 (ii) and (iii) that every $g \in \Gamma(1)$ can be expressed as a product of matrices corresponding to Type I and II reductions of forms.//

We have the relations $S^2 = (ST)^3 = I$ in $\Gamma(1)$, but it can be shown that these are essentially the only ones. In other words, $\Gamma(1)$ is the free product $\langle S \rangle * \langle ST \rangle \cong \mathbb{Z}_2 * \mathbb{Z}_3$.

Of course the essential point of Proposition 17 is that $\mathcal{F}$ is a fundamental domain for the action of $\Gamma(1)$ on $\mathbb{H}$. With a little more care (the details are a little messy) we can show how $\Gamma(1)$ identifies the sides of $\mathcal{F}$ and hence gain a description of the quotient space $\mathbb{H}/\Gamma(1)$. The translates of $\mathcal{H}$ under $\Gamma(1)$ give the famous modular diagram (see below). I have indicated how the sides of $\mathcal{F}$ are identified, and it can be seen that $\mathbb{H}/\Gamma(1)$ is topologically a sphere minus a point. We will return to this in Chapter 2. What we really wanted Proposition 17 for, however, was the information it gives about j. This is contained in the following result, which is really just a summary of what has been said before.

******see below********

Figure 1.1: The Famous Modular Diagram

Theorem 19 *$j : \mathbb{H} \to \mathbb{C}$ is invariant under the action of $SL_2(\mathbb{Z})$ on $\mathbb{H}$. In particular for all $\tau \in \mathbb{H}$ there exists $\tau' \in \mathcal{F}$ with $j(\tau) = j(\tau')$. Furthermore j is injective on the interior of $\mathcal{F}$, and identifies the boundary $\partial\mathcal{F}$ as illustrated in Figure 1.1//*

We also record the following result.

Theorem 20 *$j : \mathbb{H} \to \mathbb{C}$ is locally 1-1 on $\mathbb{H}$ except at $\Gamma(1)$-translates of the points $\rho = e^{2\pi i/3}$ and i. We have $v_\rho(j - j(\rho)) = 3$, $v_i(j - j(i)) = 2$ (i.e. j is 3-1 at ρ and its $\Gamma(1)$-translates, and 2-1 at i and its translates).*

Proof It is clear from Proposition 17 that j is locally 1-1 on $\mathcal{F}^\circ$ and its translates. The rest of the statement becomes clear upon careful consideration of how $\mathcal{F}$ maps into each of its "neighbours" on the modular diagram (Figure ????).//

Let us now finish the proof that j is surjective. Since $\mathbb{H}$ is open and $j : \mathbb{H} \to \mathbb{C}$ is analytic, it follows from the Open Mapping Theorem that $j(\mathbb{H})$ is open. For our result then, it suffices to show that $j(\mathbb{H})$ is closed. Suppose that $\{\eta_i\}_{i=1}^\infty$ is a sequence of values in $j(\mathbb{H})$ with $\eta_i \to \eta$. By Theorem 19 we can choose $\{\tau_i\}_{i=1}^\infty \subseteq \mathcal{F}$ so that $j(\tau_i) = \eta_i$. We would now like to pick a subsequence of the τ_i converging to some $\tau \in \mathcal{F}$. The continuity of j would

then give $j(\tau) = \eta$, and so $\eta \in j(\mathbb{H})$. However $\mathcal{F}$ as it stands is not compact so we cannot do this immediately. Things would go wrong if the imaginary parts $\Im(\tau_i)$ could get arbitrarily large. This, however, cannot occur.

Theorem 21 $\lim_{\Im(\tau)\to\infty} j(\tau) = \infty$.

Proof One has

$$\begin{aligned}\lim_{\Im(\tau)\to\infty} g_2(\tau) &= 120\zeta(4) + \lim_{\Im(\tau)\to\infty} \sum_{\substack{m,n \\ n\neq 0}} (m+n\tau)^{-4} \\ &= 120\zeta(4) + \sum_{\substack{m,n \\ n\neq 0}} \lim_{\Im(\tau)\to\infty} (m+n\tau)^{-4} \\ &= 120\zeta(4)\end{aligned}$$

and

$$\lim_{\Im(\tau)\to\infty} g_3(\tau) = 280\zeta(6),$$

where we have used the uniform convergence in $\mathcal{F}$ of the sums involved here. Recalling the classical evaluations $\zeta(4) = \frac{\pi^4}{90}$, $\zeta(6) = \frac{\pi^6}{945}$, one has the limits

$$\lim_{\Im(\tau)\to\infty} g_2(\tau) = \frac{4\pi^4}{3},$$

$$\lim_{\Im(\tau)\to\infty} g_3(\tau) = \frac{8\pi^6}{27} \quad \text{and}$$

$$\lim_{\Im(\tau)\to\infty} \Delta(\tau) = \left(\frac{4\pi^4}{3}\right)^3 - 27\left(\frac{8\pi^6}{27}\right)^2 = 0.$$

Therefore $\lim_{\Im(\tau)\to\infty} j(\tau) = \infty$.//

It follows that in practise the τ_i that we defined above are constained within some compact region $\mathcal{F} \cap \{z | \Im(z) \leq C\}$, and so by the argument outlined above $j(\mathbb{H})$ is closed, and we have at last

Theorem 22 *$j : \mathbb{H} \to \mathbb{C}$ is surjective.//*

As we have noted, Theorem 9 (Uniformization Theorem for Elliptic Curves over $\mathbb{C}$) follows immediately. We also have the following important Corollary.

Corollary 23 *We can give a well-defined map*

$$j : \text{Elliptic Curves over } \mathbb{C} \to \mathbb{C}$$

in the following manner. Take a Weierstraß equation $y^2 = 4x^3 - g_2x - g_3$ for E, and set $j(E) = j(g_2, g_3)$.

Proof Take two different Weierstraß representations

$$E \cong E(g_2, g_3) \cong E(g_2', g_3')$$

for E. We need to show $j(g_2, g_3) = j(g_2', g_3')$. Choose lattices Λ and Λ' with $j(g_2, g_3) = j(\Lambda) = j(g_2(\Lambda), g_3(\Lambda))$ and $j(g_2', g_3') = j(\Lambda') = j(g_2(\Lambda'), g_3(\Lambda'))$. Then

$$\begin{aligned} \mathbb{C}/\Lambda &\cong E(g_2(\Lambda), g_3(\Lambda)) \\ &\cong E(g_2, g_3) \\ &\cong E(g_2', g_3') \\ &\cong E(g_2(\Lambda'), g_3(\Lambda')) \\ &\cong \mathbb{C}/\Lambda', \end{aligned}$$

by Theorem 5 and Lemma 15. Hence, by Theorem 8, Λ and Λ' are homothetic and so have the same j-invariant. So $j(g_2, g_3) = j(g_2', g_3')$ as required.//

Finally, we can show as promised that j parametrises Elliptic Curves.

Corollary 24 $j(E) \cong j(E') \Leftrightarrow E \cong E'$.

Proof This is just Lemma 15 and Corollary 23 together.//

Let us now prove a result which will be useful later, and which offers some insight into why j should have interesting arithmetical properties. We start with some preliminaries.

1.4.3 Galois actions on Elliptic Curves

Let $\sigma \in \mathrm{Gal}(\mathbb{C}/\mathbb{Q})$. Then σ acts on the set of all Elliptic Curves over $\mathbb{C}$ by sending $E = E(g_2, g_3)$ to $E^\sigma = E(g_2^\sigma, g_3^\sigma)$. There are two crucial observations concerning this.

Observation 25 $\mathrm{End}(E) \cong \mathrm{End}(E^\sigma)$.//

Proof If $\phi : E \to E$ is an endomorphism of E then ϕ^σ (defined in the obvious way) is an endomorphism of E^σ.

Observation 26 $j(E^\sigma) = (j(E))^\sigma$.//

Proof This is immediate from the definition $j(g_2, g_3) = \frac{1728 g_2^3}{g_2^3 - 27 g_3^2}$.//

This allows us to say something about curves with complex multiplication by the ring of integers $\mathcal{O}_K$ of an imaginary quadratic field.

Theorem 27 *Suppose that E is an elliptic curve over $\mathbb{C}$ with $\mathrm{End}(E) \cong \mathcal{O}_K$. Then $j(E) \in \overline{\mathbb{Q}}$, and in fact $[\mathbb{Q}(j(E)) : \mathbb{Q}] \leq h_K$ where h_K is the class number of K.*

Proof Observation 25 says that for any $\sigma \in \mathrm{Gal}(\mathbb{C}/\mathbb{Q})$, $\mathrm{End}(E^\sigma) \cong \mathrm{End}(E)$ $\cong \mathcal{O}_K$. But Theorem 14 tells us that there are exactly h_K different isomorphism classes of curve with endomorhism ring $\mathcal{O}_K$. Since j parametrises isomorphism classes of curves (Corollary 24) this means that $j(E^\sigma) = (j(E))^\sigma$. So $j(E)$ has at most h_K $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$-conjugates, and by standard Field Theory $[\mathbb{Q}(j(E)) : \mathbb{Q}] \leq h_K$.//

Before moving on to matters modular, I find it impossible to resist giving the following classical application of the machinery we have developed in this first chapter.

Theorem 28 (Picard's Little Theorem) *An entire function whose range omits two values is constant.*

Proof We may assume the omitted values are 0 and 1, so that we have an analytic function $f : \mathbb{C} \to \mathbb{C} \setminus \{0, 1\}$. From Theorems 22 and 20 we see that j restricts to an *unbranched covering* $j : \hat{\mathbb{H}} \to \mathbb{C} \setminus \{j(\rho), j(i)\}$ where $\hat{\mathbb{H}}$ is the half-plane $\mathbb{H}$ with the $\Gamma(1)$-translates of i and ρ removed. Clearly we can use this to produce an unbranched covering $\gamma : \hat{\mathbb{H}} \to \mathbb{C} \setminus \{0, 1\}$. Now in such a situation we can find a *lift* $\tilde{f}$ of f. Such a function maps $\mathbb{C}$ to $\hat{\mathbb{H}}$, is analytic, and makes the following commute:

$$\begin{array}{ccc} \mathbb{C} & \xrightarrow{\tilde{f}} & \hat{\mathbb{H}} \\ & \searrow & \downarrow \gamma \\ & & \mathbb{C} \setminus \{0, 1\} \end{array}$$

But the map $\theta : z \to \frac{1}{z+i}$ sends $\hat{\mathbb{H}}$ into a bounded domain. Hence $\theta \circ \tilde{f} : \mathbb{C} \to \mathbb{C}$ is bounded and so, by Liouville's Theorem, must be constant. It is easy to deduce now that f must be constant.//

Chapter 2

A the end of Chapter 1 we proved (Theorem 27) that the j-invariant of an elliptic curve over $\mathbb{C}$ with complex multiplication by $\mathcal{O}_K$ is an algebraic number. This however was simply a taster for the end result of this chapter, which we now state.

Theorem 29 *Let E be any elliptic curve with complex multiplication. Then $j(E)$ is an algebraic integer.*

This is clearly an unexpected and remarkable result. There are many proofs, all reasonably non-trivial. Silverman [15] presents three of these. We present an expanded version of the first of Silverman's proofs. The other two use much deeper theory of Elliptic Curves.

Let us begin with an attempt to motivate the proof. Suppose that $E \cong \mathbb{C}/\Lambda$ with $\Lambda = \mathbb{Z} \oplus \mathbb{Z}\eta$, $\eta \notin \mathbb{R}$. We showed in Consequence 9.2 that E has complex multiplication if and only if η is a quadratic irrationality. In this case there is a non-vacuous relation

$$\eta = \frac{a\eta + b}{c\eta + d} \tag{2.1}$$

with $a, b, c, d \in \mathbb{Z}$. We write this as $\eta = M\eta$, where $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Clearly, then, $j(\eta) = j(M\eta)$. Suppose we step back from this specific η and ask ourselves whether there is a general relation between $j(\tau)$ and $j(M\tau)$, valid for all $\tau \in \mathbb{H}$. Substituting $\tau = \eta$ in such a relation might then give us some information about $j(\eta)$. In view of our goal, Theorem 29, we would like to show in fact that there is a polynomial P such that $P(j(\tau), j(M\tau)) = 0$ for all $\tau \in \mathbb{H}$. This would give $P(j(\eta), j(\eta)) = 0$ and we would hope to be able to derive from this a monic polynomial over $\mathbb{Z}$ satisfied by $j(\eta)$.

This is exactly the approach we use. Our first objective is to look at why such a P should exist. Along the way we will derive results which will help us construct P fairly explicitly. This will allow us to prove Theorem 29 as outlined above.

2.1 Modular Groups, Curves and Functions

2.1.1 $X(1)$

Thus far we have not really expanded on the properties of j described in Theorems 19, 21 and 22. Firstly we discuss the idea that $\mathbb{H}/\Gamma(1)$ is topologically a sphere minus a point, being $\mathcal{F}$ with its sides identified as in Figure 1.1. One aim is to add in the extra point. This is clearly going to arise from the "point at ∞ of $\mathbb{H}$". However $\mathbb{H} \cup \infty$ is not invariant under $\Gamma(1)$, so we must also add in the $\Gamma(1)$-orbit of ∞. What we get is the set of rational points on the real line, and we denote this together with ∞ by $\mathbb{P}^1(\mathbb{Q})$. we write $\mathbb{H}^* = \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ and call it the *extended upper half-plane*. The images of $\mathbb{P}^1(\mathbb{Q})$ under the action of $\Gamma(1)$ is called the *cusp* of $\Gamma(1)$. So now we have an action of $\Gamma(1)$ on $\mathbb{H}^*$ and we feel that $\mathbb{H}^*/\Gamma(1)$ should be topologically a sphere. Of course this makes no sense as we have not defined a topology on $\mathbb{H}^*/\Gamma(1)$. What we want to do is define a topology on $\mathbb{H}^*$ and then give $\mathbb{H}^*/\Gamma(1)$ the quotient topology. A little thought shows that the correct way to do this is to take our open sets to be

(1) The open sets of $\mathbb{H}$
(2) The sets $S_C = \{\tau | \Im(\tau) > C\}$ (the "neighbourhoods of ∞") and
(3) The $\Gamma(1)$-translates of the S_C, i.e. all open discs tangent to the real line at a point of $\mathbb{Q}$.

This sphere will be called $X(1)$, but before we may permit ourselves to name it thus we must give it the structure of a Riemann Surface. It is no surprise that j has a rôle to play in all this, and Theorems 19, 21 and 22 can be more or less summed up in the statement that $j : X(1) \to \mathbb{C}_\infty$ is a bijection (This is a slight abuse of notation; we really mean $(\pi^{-1})^* j$, where $\pi : \mathbb{H}^* \to X(1)$ is projection. But this is a touch pedantic). This bijection allows us to define a complex structure on $X(1)$ by pulling back that on $\mathbb{C}_\infty$. Of course $X(1)$ is then isomorphic to $\mathbb{C}_\infty$ as a Riemann Surface. Also $\pi : \mathbb{H} \to X(1)$ is analytic. [1]

Now associated with any Riemann Surface are its meromorphic functions. The functions on $\mathbb{H}^*$ corresponding to meromorphic functions on $X(1)$ are of some interest and are described in Theorem 31. First, however, we need

[1] This Riemann Surface can also be constructed from scratch geometrically, without the need for j. In Jones and Singerman [7] the construction of the general quotient surface $\mathbb{H}^*/G$ is given, where G is any Fuchsian Group (discrete subgroup of $PSL_2(\mathbb{R})$). the construction is essentially determined by the requirement that $\pi : \mathbb{H} \to \mathbb{H}^*/G$ be analytic,a s well as requiring "good behaviour" at the cusps. It is worthwhile to mention that slight complications occur at "elliptic points" $P \in \mathbb{H}^*$, where π is not locally 1-1. For example it can be shown that $\rho = e^{2\pi i/3}$ has stabilizer of order 3 under the action of $\Gamma(1)$, and so $\pi : \mathbb{H}^* \to \mathbb{H}^*/\Gamma(1)$ is 3-1 about ρ. Hence a local parameter at $\pi(\rho)$ must lift to a function which is locally 3-1 at ρ. j does indeed have this property, by Theorem 20.

to introduce some new concepts.

Definition 30 *Let $f : \mathbb{H} \to \mathbb{C}_\infty$ be meromorphic, $\Gamma(1)$-invariant and holomorphic in some neighbourhood of ∞. Then $f(\tau + 1) = f(\tau)$ for all τ and so f has a* q-expansion

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

in terms of $q = e^{2\pi i \tau}$. this is really just a Laurent Series, and is valid in some neighbourhood $0 < |q| < C$ of ∞ (note that $q \to 0$ corresponds to $\Im(\tau) \to \infty$). We say f is meromorphic at ∞ *if there exists N such that $a_n = 0$ for all $n < -N$. The least such N is the order of the pole at ∞ of f. When $N = 0$ we say that f is* holomorphic at ∞ *and in this case we can meaningfully define $f(\infty) = a_0$.*

By standard complex analysis (Riemann's Removable Singularity Theorem and Taylor's Theorem) f is meromorhic at ∞ if and only if $\lim_{\Im(\tau)\to\infty} f(\tau)$ exists. If this limit is finite, f is holomorhic at ∞. it follows from Theorem 21 that $j : \mathbb{H} \to \mathbb{C}_\infty$ is meromorphic at ∞.

Theorem 31 *The meromorphic functions on $X(1)$ correspond to functions $f : \mathbb{H}^* \to \mathbb{C}$ satisfying*

(1) *f is $\Gamma(1)$-invariant*
(2) *f is meromorphic on $\mathbb{H}$ and at ∞.*

That is to say, f is meromorphic on $\mathbb{H}^*$.

Proof (1) is obvious. The fact that meromorphic functions on $X(1)$ correspond to meromorphic functions on $\mathbb{H}^*$ can be checked in detail using the charts on $X(1)$ and Definition 30 if desired. Conceptually, however, the crucial fact is that j defines the complex structure on $X(1)$ and j is meromorphic on $\mathbb{H}^*$.//

Functions satisfying the conditions of Theorem 31 form a field $\mathbb{C}(X(1))$ and are deemed to be $\Gamma(1)$-modular of weight 0. Of course, the j-function is such a function. In fact we know from elementary Riemann Surface Theory that the meromorphic functions on $\mathbb{C}_\infty$ are just the rational functions. It is therefore easy to deduce

Theorem 32 *The field of $\Gamma(1)$-modular functions of weight 0 on $\mathbb{H}^*$ is just $\mathbb{C}(j)$.*

Things become really interesting when we look at differential k-forms on $X(1)$. These give so-called modular functions and modular forms of weight $2k$, which have a bewildering array of fascinating properties. Some of these will be discussed in Chapter 3. In this chapter "modular function" will always mean "modular function of weight 0".

2.1.2 Higher Modular Groups and Curves

Now recall that we were interested in studying the function

$$f_M : \tau \to j(M\tau)$$

on $\mathbb{H}$, and in finding a relationship between it and $j(\tau)$. Here $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is any invertible matrix with $a, b, c, d \in \mathbb{Z}$ and $ad - bc = n$, and we take w.l.o.g. $\gcd(a, b, c, d) = 1$. Of course if f_M were $\Gamma(1)$-modular, we would be done. However although f_M satisfies the second condition of Theorem 31, it need not be $\Gamma(1)$-invariant. Indeed suppose f is γ-invariant where $\gamma \in \Gamma(1)$. Then

$$\begin{array}{ll} & f_M(\tau) = f_M(\gamma\tau) \qquad (\forall \tau \in \mathbb{H}) \\ \Leftrightarrow & j(M\tau) = j(M\gamma\tau) \qquad (\forall \tau \in \mathbb{H}) \\ \Leftrightarrow & M\tau = gM\gamma\tau \qquad \text{for some } g \in \Gamma(1) \text{ and all } \tau \in \mathbb{H} \\ \Leftrightarrow & \gamma \in M^{-1}\Gamma(1)M \cap \Gamma(1). \end{array}$$

Now the set of such γ is clearly a subgroup of $\Gamma(1)$, which we call $G(M)$. It would be of interest to describe this in more detail. there are two important points to note:

Fact 1. If $M' = \gamma M$ for some $\gamma \in \Gamma(1)$ then $f_M = f_{M'}$ and $G(M) = G(\gamma M)$.

Fact 2. If $M' = M\gamma$ for some $\gamma \in \Gamma(1)$ then $G(M') = \gamma^{-1}G(M)\gamma$ is conjugate to $G(M)$.

This suggests we look not just at M but at all M' of the form $\gamma_1 M \gamma_2$ for $\gamma_1, \gamma_2 \in \Gamma(1)$. There are lots of such M', but they have two common features- they all have determinant n, and their entries have no non-trivial common factor. Let us, then, consider the right and left actions of $\Gamma(1)$ on the whole set $\mathcal{S}_n$ of matrices $\begin{pmatrix} i & j \\ k & l \end{pmatrix}$ with $il - jk = n$ and $\gcd(i, j, k, l) = 1$.

Theorem 33 (Left action of $\Gamma(1)$ on $\mathcal{S}_n$) *For every $A \in \mathcal{S}_n$ there is $\gamma \in \Gamma(1)$ such that $\begin{pmatrix} i' & j' \\ k' & l' \end{pmatrix} = \gamma A$ is "reduced", i.e. $k' = 0$, $i'l' = n$, $0 \le j' < l'$ and $\gcd(i', j', k', l') = 1$. Furthermore these reduced matrices are mutually inequivalent under the left action of $\Gamma(1)$.*

Proof It is easy to find γ. First get $k' = 0$, and then get j' in the required range using $\gamma = \begin{pmatrix} 1 & \lambda \\ 0 & l \end{pmatrix}$. We will then automatically have $i'l' = n$ and $\gcd(i', j', k', l') = 1$. The proof that reduced matrices are inequivalent is straightforward and unenlightening.//

We denote the set of reduced matrices in $\mathcal{S}_n$ by $\mathcal{R}_n$.

Theorem 34 (Right action of $\Gamma(1)$ on $\mathcal{S}_n$) *The right action of $\Gamma(1)$ on $\mathcal{S}_n$ is transitive on the orbits under the left action described above. In other words any two matrices $A, A' \in \mathcal{S}_n$ are related by $A' = \gamma_1 A \gamma_2$ for some $\gamma_1, \gamma_2 \in \Gamma(1)$.*

Proof Write $A \sim A'$ for an equivalence of the above kind. It suffices to show that $A \sim P_n = \begin{pmatrix} n & 0 \\ 0 & l \end{pmatrix}$ for all A. This is another unenlightening fiddle with explicit matrices.//

It follows from Facts 1 and 2 together with Theorems 33 and 34 that the groups $G(A)$ ($A \in \mathcal{S}_n$) are precisely the conjugates in $\Gamma(1)$ of the group $G(P_n)$. However it is a triviality to verify that

$$G(P_n) = \left\{ \begin{pmatrix} p & q \\ r & s \end{pmatrix} \in SL_2(\mathbb{Z}) \;\middle|\; q \equiv 0 \pmod{n} \right\}$$

It is standard to call this group $\Gamma_0(n)$. We can summarise much of the above in the following

Proposition 35 *The function $f_M : \tau \to j(M\tau)$ is $G(M)$-modular (of weight 0). $G(M) = M^{-1}\Gamma(1)M \cap \Gamma(1)$ is a conjugate of $\Gamma_0(n)$ in $\Gamma(1)$.//*

Note, however, that we have not fully defined what it means for a function to be $G(M)$-modular. The problem is that there may well be more than one cusp in $\mathbb{H}^*/G(M)$- that is to say $G(M)$ need not act transitively on $\mathbb{P}^1(\mathbb{Q})$ as was the case for $\Gamma(1)$. For example the action of $\Gamma_0(p)$ on $\mathbb{P}^1(\mathbb{Q})$, where p is a prime, gives rise to the two orbits

$$\begin{array}{rcll} S_1 & = & \left\{ \frac{q}{r} \;\middle|\; \gcd(q,r) = 1, q \not\equiv 0 \pmod{p} \right\} \cup \{\infty\} & \text{and} \\ S_2 & = & \left\{ \frac{q}{r} \;\middle|\; \gcd(q,r) = 1, q \equiv 0 \pmod{p} \right\} & \end{array}$$

There will however only be finitely many cusps on $\mathbb{H}^*/G(M)$. In particular they are isolated, and so we deem a $G(M)$-invariant function $f : \mathbb{H} \to \mathbb{C}_\infty$ to be meromorphic at a cusp if it is continuous there (cf. Riemann's Removable Singularity Theorem). Shimura [12] can be consulted for more details. One then defines a $G(M)$-modular function of weight 0 to be one satisfying

(i) $f : \mathbb{H} \to \mathbb{C}_\infty$ is $G(M)$-invariant
(ii) f is meromorphic at *all* the cusps on $\mathbb{H}^*/G(M)$.

Recall our first goal was to motivate the existence of a polynomial $P \in \mathbb{C}[X, Y]$ such that $P(j(\tau), j(M\tau)) = 0$ for all $\tau \in \mathbb{H}$. this we can now do. First of all we observe that j itself is a $G(M)$-modular function. It is certainly $G(M)$-invariant because $G(M) \subseteq \Gamma(1)$. Now we obtained information about the $\Gamma(1)$-modular functions by looking at $\mathbb{H}^*/\Gamma(1)$ as a Riemann Surface. To describe $G(M)$-modular functions then, we look at $\mathbb{H}^*/G(M)$. This

can be given the natural structure of a compact connected Riemann Surface $X(M)$ such that $\pi_M : \mathbb{H}^* \to X(M)$ is analytic (see Shimura [12]).

Now if M and M' are in the same left $\Gamma(1)$-orbit in $\mathcal{S}_n$ then we know that $G(M) = G(M')$ and so $X(M) = X(M')$. It turns out that in fact all the $X(M)$ ($M \in \mathcal{S}_n$) are isomorphic. To prove this it suffices, by Theorem 34, to consider the case $M' = M\gamma$. Then $G(M') = \gamma^{-1}G(M)\gamma$ and so

$$\tau \overset{G(M')}{\sim} \tau' \iff \gamma\tau \overset{G(M)}{\sim} \gamma\tau'$$

It follows that the map $\gamma : \mathbb{H}^* \to \mathbb{H}^*$ factors through the projections $\pi_M, \pi_{M'}$ to give an isomorphism $\iota : X(M') \to X(M)$. The situation is illustrated in the following diagram.

$$\begin{array}{ccc} \mathbb{H}^* & \xrightarrow{\gamma} & \mathbb{H}^* \\ \downarrow{\scriptstyle \pi_{M'}} & & \downarrow{\scriptstyle \pi_M} \\ X_{M'} & \xrightarrow{\iota} & X_M \end{array}$$

In particular, all the $X(M)$ are isomorphic to $X_0(n) = \mathbb{H}^*/\Gamma_0(n)$. Now (exactly as with $X(1)$) the meromorphic functions on $X(M)$ correspond to $G(M)$-modular functions of weight 0. $j(\tau)$ and $j(M\tau)$ are both examples of such functions. But it is shown in any suitably advanced text on Riemann Surfaces that any two meromorphic functions on a compact Riemann Surface satisfy some polynomial over $\mathbb{C}$ [2]

Hence we do indeed get a relation $P(j(\tau), j(M\tau)) = 0$ for some $P \in \mathbb{C}[X, Y]$ and all $\tau \in \mathbb{H}$.

We can also guess as to what P must be. Suppose $M' = M\gamma$. Then $\gamma^* j(\tau) = j(\gamma\tau) = j(\tau)$ and $\gamma^* j(M\tau) = j(M\gamma\tau) = j(M'\tau)$. In words, the function $j(M\tau)$ on $X(M)$ "corresponds" to $j(M'\tau)$ on $X(M')$ under ι^* and j corresponds to itself. Since ι^* is an isomorphism of function fields, we have $P(j(\tau), j(M\tau)) = 0 \Leftrightarrow P(j(\tau), j(M'\tau)) = 0$. This suggests in view of Theorem 34 that the P we are looking for is not specific to M, but is dependant only on $\det M = n$. This turns out to be the case, and the polynomial in question is called the *modular equation* of level n. We denote it by Φ_n, and will construct it below. We note however that an explicit determination of the coefficients of Φ_n, even for $n = 2$, is highly non-trivial. Φ_2 is given in Silverman [15].

Before constructing Φ_n, however, we pause to discuss a few aspects of these

[2] In fact this result is not hard to prove. And then, once we know that there *are* non-trivial meromorphic functions on R, it follows that $\mathbb{C}(R)$ has transcendence degree 1. This can be used to deduce that Riemann Surfaces are essentially the same thing as algebraic curves.

******PICTURE!!!!*****

Figure 2.1: The Modular Curve $X_0(2)$

new surfaces $X_0(n)$ that have arisen in this discussion. One should perhaps keep in mind a picture of $X_0(n)$ as a certain number of translates of $\mathcal{F}$, corresponding to cosets of $\Gamma_0(n)$ in $\Gamma(1)$, together with certain identifications. For example we have Figure 2.1, which shows $X_0(2)$ to be topologically a sphere: Perhaps the most obvious question to ask is "What is the genus of $X_0(n)$?" This can be calculated explicitly using the Riemann-Hurwitz Relation. The key observation is that, since $\Gamma_0(n) \leq \Gamma(1)$, we have a branched covering

$$\pi : X_0(n) \to X(1) \cong \mathbb{C}_\infty$$

The number of sheets, $\deg(\pi)$, is equal to the index $[\Gamma(1) : \Gamma_0(n)]$. By exhibiting a bijection of sets

$$\theta : \Gamma(1)/\Gamma_0(n) \to \mathcal{R}_n,$$

where $\mathcal{R}_n$ is the set of "reduced" matrices defined earlier, one can show that

$$\deg(\pi) = n \prod_{p|n} \left(1 + \frac{1}{p}\right)$$

It is not hard to see that ramification can only occur above the points $\rho = e^{2\pi i/3}$, i and ∞ of $X(1)$, and a detailed analysis of the behaviour at these

points enables one to deduce $g(X_0(n))$. The formula is complicated in its full generality (see, for example, Hüsemoller [6]) but simplifies considerably when $n = p$ is a prime. Indeed one has $g(X_0(2)) = g(X_0(3)) = 0$,

$$\begin{aligned} g(X_0(p)) &= \tfrac{p+1}{12} - \tfrac{1}{4}\left(1 + \left(\tfrac{-1}{p}\right)\right) - \tfrac{1}{3}\left(1 + \left(\tfrac{-3}{p}\right)\right) \quad (p \geq 5) \\ &= \begin{cases} \lfloor \frac{p+1}{12} \rfloor - 1 & (p \equiv 1 (\mathrm{mod} 12)) \\ \lfloor \frac{p+1}{12} \rfloor & (\text{otherwise}) \end{cases} \end{aligned}$$

In particular the only p for which $g(X_0(p))$ is zero are $p = 2, 3, 5, 7, 13$. For those p the $\Gamma_0(p)$-modular functions of weight 0 form a rational function field $\mathbb{C}(t)$. t is called an *absolute invariant* of $\Gamma_0(p)$; Theorem 32 then says that j is an absolute invariant for $\Gamma(1)$.

For $p = 11, 17, 19, 23$, $X_0(p)$ is an elliptic curve. For example $X_0(11)$ is isomorphic to the curve $y^2 = x(x^3 - 20x^2 + 56x - 44)$ (see McKean and Moll [8]- note that this curve is not in Weierstraß Form). This gives much extra structure to $X_0(11)$, which may be used to study the arithmetic of the curve. The famous Taniyama-Weil conjecture states in part that we can study *any* elliptic curve over $\mathbb{Q}$ using higher modular curves, in the following sense-

Conjecture 36 (Taniyama-Weil, weak form) *Let E be an elliptic curve defined over $\mathbb{Q}$. Then for some n there is a surjective map $\phi : X_0(n) \to E$ defined over $\mathbb{Q}$.*

A partial solution of this conjecture was used by Andrew Wiles in his proof of Fermat's Last Theorem. We conclude this digression by mentioning that there are two other important families $\{\Gamma(n) | n \in \mathbb{N}\}$ and $\{\Gamma_1(n) | n \in \mathbb{N}\}$ of "congruence subgroups" of $\Gamma(1)$which find application in number theory. These are defined by

$$\begin{aligned} \Gamma_1(n) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1) \;\middle|\; c \equiv 0 (\mathrm{mod}\ n), a \equiv d \equiv 1 (\mathrm{mod}\ n) \right\} \\ \Gamma(n) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1) \;\middle|\; a \equiv d \equiv 1 (\mathrm{mod}\ n), b \equiv c \equiv 0 (\mathrm{mod}\ n) \right\} \end{aligned}$$

One can consider modular functions for these groups, and also the quotient surfaces $X_1(n)$ and $X(n)$. See Silverman [14], [15] for more details.

2.2 The Modular Equation, and the Integrality of the j-Invariant

We now take a more concrete approach. Without further ado, define

$$\begin{aligned}\Psi_n(X;\tau) &= \prod_{M\in\mathcal{R}_n}(X - j(M\tau)) \\ &= \prod_{\substack{ad=n\\ \gcd(a,b,d)=1\\ 0\le b<d}} \left(X - j\left(\frac{a\tau+b}{d}\right)\right) \end{aligned} \tag{2.2}$$

The product picks out exactly one M from each orbit in the left action of $\Gamma(1)$ on $\mathcal{S}_n$, by Theorem 33. Certainly then $\Psi_n(j(M\tau),\tau) = 0$ for any $M \in \mathcal{S}_n$. Our aim is to show that $\Psi_n(X;\tau)$ has the form $\Phi_n(X, j(\tau))$ for a polynomial Φ_n. Now the coefficient $a_i(\tau)$ of X^i is a symmetric function in the $j(M\tau)$, and so is $\Gamma(1)$-invariant as a function of τ. Indeed if $M_1,\ldots,M_N$ are a distinct set of representatives for the orbits of the left $\Gamma(1)$-action on $\mathcal{S}_n$, then so are $M_1\gamma,\ldots M_N\gamma$ for any $\gamma\in\Gamma(1)$. Hence $\{j(M_1\gamma\tau),\ldots,j(M_N\gamma\tau)\}$ is a permutation of $\{j(M_1\tau),\ldots,j(M_N\tau)\}$. But a_i is meromorphic on $\mathbb{H}^*$, and so is in fact $\Gamma(1)$-modular. It is therefore a rational function of j by Theorem 32. However none of the functions $j\left(\frac{a\tau+b}{d}\right)$ has a pole in $\mathbb{H}$ and hence neither does a_i. It follows that a_i is actually a *polynomial* in j. So indeed we have a polynomial $\Phi_n(X,Y)$ with $\Phi_n(j(M\tau),j(\tau)) = 0$ for all $\tau\in\mathbb{H}$ and any $M\in\mathcal{S}_n$. it is defined by

$$\Phi_n(X, j(\tau)) = \prod_{\substack{ad=n\\ \gcd(a,b,d)=1\\ 0\le b<d}} \left(X - j\left(\frac{a\tau+b}{d}\right)\right) \tag{2.3}$$

As we remarked earlier, the fact that $E = \mathbb{Z}\oplus\mathbb{Z}\eta$ has complex multiplication gives a non-trivial relation $\eta = M\eta$ ($M\in\mathcal{S}_n$), and this exhibits $j(\eta)$ as a solution of the diagonal equation $\Phi_n(j(\eta),j(\eta)) = 0$.

Our next objective, then, is to start proving things about the coefficients of Φ_n so as to enable us to deduce Theorem 29. However there is a very good reason why we cannot do this immediately. For Theorem 29 is clearly very dependant on the factor 1728 in the definition of j, and we have yet to use this in any way. in fact we could have got this far with any non-zero rational number in place of 1728. The reson for the choice is the following remarkable result.

Theorem 37 *The q-expansion of j (cf Definition 30) is*

$$\begin{aligned} j(\tau) &= q^{-1} + 744 + 196884q + 21493760q^2 + \ldots \\ &= q^{-1} + \textstyle\sum_{n=0}^{\infty} c(n)q^n \end{aligned}$$

where the $c(n)$ are integers.

We will prove Theorem 37 in the next Chapter, but our immediate objective is to press on with the proof of Theorem 29. The next result is the following.

Theorem 38 $\Phi_n(X,Y) \in \mathbb{Z}[X,Y]$.

Proof It clearly suffices to show that $a_i \in \mathbb{Z}[j]$, where a_i is any symmetric function in the elements of $\{j(M\tau) \mid M \in \mathcal{R}_n\}$ - see "construction of Φ_n". Now (as we remarked) a_i is $\Gamma(1)$-modular, and so has a q-expansion. In view of Theorem 37, it is this we look at first.

Lemma 39 *The q-expansion of $a_i(\tau)$ lies in $\mathbb{Z}[\![q,q^{-1}]\!]$.*

Proof of Lemma In the following set $\zeta = e^{2\pi i/n}$ and $Q = q^{1/n} = e^{2\pi i\tau/n}$. Then it is immediate to check that if $M = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in \mathcal{S}_n$ (so that $ad = n$) then

$$j(M\tau) = \zeta^{-ab}Q^{-a^2} + \sum_{k=0}^{\infty} c(k)\zeta^{abk}Q^{a^2k} \tag{2.4}$$

in some annulus $0 < Q < c_M$. By standard analysis (see Rudin [10]) we may add and multiply these series for various M term-by-term to get a series expansion for a_i in $\mathbb{Z}[\zeta][\![Q,Q^{-1}]\!]$ and valid in some annulus $0 < Q < c$. However $a_i(\tau)$ is $\Gamma(1)$-modular, and so has 1 as a period. it follows that this series for $a_i(\tau)$ actually lies in $\mathbb{Z}[\zeta][\![q,q^{-1}]\!]$. However we are to show that this series for $a_i(\tau)$does in fact lie in $\mathbb{Z}[\![q,q^{-1}]\!]$. Although this is just a combinatorial statement about "how the powers of ζ cancel" it is most naturally proved in the language of Galois Theory. Let $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$. Then $\sigma(\zeta) = \zeta^r$ for some $r = r(\sigma)$ with $\gcd(r,n) = 1$, because $\sigma(\zeta)$ must have order exactly n in $\mathbb{C}^*$. Let σ act on the Q-expansion of $j(M\tau)$, coefficient by coefficient. Here Q is being regarded as an indeterminate. We get

$$\begin{aligned} (j(M\tau))^\sigma &= \zeta^{-rab}Q^{-a^2} + \textstyle\sum_{k=0}^{\infty} c(k)\zeta^{abrk}Q^{a^2k} \\ &= j(M'\tau) \text{ where } M' = \begin{pmatrix} a & rb \\ 0 & d \end{pmatrix} \\ &= j(M''\tau) \text{ where } M'' = \begin{pmatrix} a & rb(\mathrm{mod}\, d) \\ 0 & d \end{pmatrix}, \end{aligned} \tag{2.5}$$

since $M' \sim M''$ under the left action of $\Gamma(1)$ on $\mathcal{S}_n$. Indeed

$$\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & \beta \\ 0 & d \end{pmatrix} = \begin{pmatrix} a & \beta + \lambda d \\ 0 & d \end{pmatrix}$$

for any λ, β. We now claim

Claim 40 *Sending M to M'' induces a permutation of the reduced matrices $\mathcal{R}_n$.*

Proof of Claim This amounts to showing that as b runs through the complete set of allowable values (i.e. $0 \leq b < d$, $\gcd(a, b, d) = 1$) then so does $rb (\mathrm{mod}\, d)$. Now $\gcd(r, n) = 1 \Rightarrow \gcd(r, d) = 1$ since $ad = n$. therefore $b \mapsto rb (\mathrm{mod}\, d)$ induces a permutation on the complete set $\{0, 1, \ldots, d-1\}$ of reduced residues. But it is easy to see that $\gcd(a, b, d) = 1 \Rightarrow \gcd(a, rb (\mathrm{mod}\, d), d) = 1$, again using the coprimality of r and d. this proves the claim.//

Claim 40 and Equation 2.5 show that if a_i is any symmetric function in the elements of $\{j(M\tau) \mid M \in \mathcal{R}_n\}$ then all coefficients γ_{ij} in the Q-expansion of a_i are $\mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$-invariant. Therefore $\gamma_{ij} \in \mathbb{Q}$. But we have already shown that $\gamma_{ij} \in \mathbb{Z}[\zeta]$; and hence $\gamma_{ij} \in \mathbb{Q} \cap \mathbb{Z}[\zeta] = \mathbb{Z}$. It follows that the q-expansion of $a_i(\tau)$ is in $\mathbb{Z}[\![q, q^{-1}]\!]$, thus completing the proof of Lemma 39.//

Now Theorem 38 is concerned with the expression for $a_i(\tau)$ as a polynomial $a_{i0} + a_{i1}j + \ldots + a_{im}j^m$ in j. Using Theorem 37 we can see that the q-expansion of $a_i(\tau)$ must be

$$a_{im}q^{-m} + \{\text{terms in } q^{-t} \text{ with } t < m\}$$

hence by Lemma 39 we have $a_{im} \in \mathbb{Z}$. Now the q-expansion of j^m can be obtained simply by formally raising the q-expansion of j to the power m, and hence lies in $\mathbb{Z}[\![q, q^{-1}]\!]$. It follows that

$$a_i(\tau) - a_{im}j^m = a_{i0} + a_{i1}j + \ldots + a_{i,m-1}j^{m-1} \in \mathbb{Z}[\![q, q^{-1}]\!].$$

We can now proceed inductively to get $a_{i,m-1} \in \mathbb{Z}$ and so on. Therefore $a_i(\tau) \in \mathbb{Z}[j]$ and Theorem 38 follows immediately.//

Finally we are ready to look at the diagonal equation $\Phi_n(X, X)$ satisfied by $j(\eta)$ in cases of interest. Theorem 38 says that this lies in $\mathbb{Z}[X]$, and we would like to show that it is in fact monic. The next Theorem shows that this is true with one small proviso.

Theorem 41 *Suppose $n > 1$ is not a square. then $\Phi_n(X, X)$ has leading coefficient ± 1.*

Proof As we saw during the proof of Theorem 38, the leading coefficient in j of $\Phi_n(j, j)$ is just the leading coefficient of the q-expansion of $\Phi_n(j, j)$- that is to say the coefficient of the most negative power of q. But Equation 2.3 gives

$$\Phi_n(j, j) = \prod_{M \in \mathcal{R}_n} (j - j \circ M) \tag{2.6}$$

Now (2.4) together with Theorem 37 shows that the Q-expansion of $j(\tau) - j(M\tau)$ is

$$\left(\frac{1}{Q^n} + \sum_{k=0}^{\infty} c(k)Q^{n/k}\right) - \left(\frac{1}{\zeta^{ab}Q^{a^2}} + \sum_{k=0}^{\infty} c(k)\zeta^{abk}Q^{a^2k}\right)$$

This has leading coefficient either 1 or ζ^{ab}; for since n is not a square, the terms Q^{-n} and $\zeta^{-ab}Q^{-a^2}$ cannot cancel. In both cases, the leading coefficient is a root of unity. Hence when multiplying all these together in the product (2.6), we get a Q-expansion for $\Phi_n(j,j)$ with some root of unity ε as the leading coefficient. However this Q-expansion is really a q-expansion (that is to say all terms Q^i with $i \not\equiv 0 (\mathrm{mod}\, n)$ vanish) with integer coefficients. Therefore $\varepsilon = \pm 1$, and we have proved Theorem 41.//

Let us assess the situation. If $\eta = M\eta$ ($M \in \mathcal{R}_n$) then $j(\eta)$ satisfies the diagonal equation $\Phi_n(j(\eta), j(\eta)) = 0$. Furthermore if n is not a square, this will be a monic polynomial in $\mathbb{Z}[X]$ and $j(\eta)$ will be an algebraic integer. We are, therefore, a short step away from proving Theorem 29.

Proof of Theorem 23 It suffices to show that we can choose n not a square. Suppose $p\eta^2 + q\eta + r = 0$ with $\gcd(p,q,r) = 1$ and $q^2 - 4pr < 0$. then it suffices for $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ to satisfy $b = -r$, $c = p$, $d - a = q$; we will have $\gcd(a,b,c,d) = 1$ automatically since $\gcd(p,q,r) = 1$. For any such set of integers we have $\det(M) = ad - bc = a^2 + qa + rp$. If this is a square, say λ^2, then $(2a+q)^2 - (2\lambda)^2 = \Delta = q^2 - 4pr$. But this can only happen for finitely many a. Therefore one can find a non-square n and a matrix $M \in \mathcal{S}_n$ with $\eta = M\eta$. By above comments, this proves the result.//

Corollary 42 *Suppose E is an elliptic curve such that* $\mathrm{End}(E) \cong \mathcal{O}_K$, *where K is a quadratic imaginary field with class number 1. then $j(E) \in \mathbb{Z}$.*

Proof This is merely a combination of Theorems 29 and 27.//

It is well-known that there are only nine such fields: they are $K = \mathbb{Q}(\sqrt{-1})$, $\mathbb{Q}(\sqrt{-2})$, $\mathbb{Q}(\sqrt{-3})$, $\mathbb{Q}(\sqrt{-7})$, $\mathbb{Q}(\sqrt{-11})$, $\mathbb{Q}(\sqrt{-19})$, $\mathbb{Q}(\sqrt{-43})$, $\mathbb{Q}(\sqrt{-67})$ and $\mathbb{Q}(\sqrt{-163})$. By the results leading up to Theorem 14 there is for each K a unique isomorphism class of elliptic curve with $\mathrm{End}(E) \cong \mathcal{O}_K$, and $E \cong \mathbb{C}/\Lambda$ for any ideal $\Lambda \triangleleft \mathcal{O}_K$. In particular we can take $\Lambda = \mathcal{O}_K$, in which case $\Lambda = \mathbb{Z} \oplus \mathbb{Z}\eta_K$ for suitable η_K. In fact for $K = \mathbb{Q}(\sqrt{-D})$, D squarefree, we take

$$\eta_K = \begin{cases} \sqrt{-D} & D \equiv 1 \text{ or } 2 \pmod 4 \\ \frac{1+\sqrt{-D}}{2} & D \equiv 3 \pmod 4 \end{cases}$$

For the nine special fields K, $j(\eta_K)$ will then be an integer. In particular, $j\left(\frac{1+\sqrt{-163}}{2}\right) \in \mathbb{Z}$. The corresponding value of $q = e^{2\pi i\tau}$ is $e^{-\pi\sqrt{163}}$, and if

we substitute this into the q-expansion for j (Theorem 37 we get

$$j\left(\frac{1+\sqrt{-163}}{2}\right) = -e^{\pi\sqrt{163}} + 744 + \sum_{n=1}^{\infty} c(n)(-1)^{n+1}e^{-n\pi\sqrt{163}} \qquad (2.7)$$

Now $e^{-\pi\sqrt{163}} < 4 \times 10^{-18}$, so provided the $c(n)$ decay "quite quickly" the sum in the above equation should be negligable. Hence

$$e^{\pi\sqrt{163}} \approx 744 - j\left(\frac{1+\sqrt{-163}}{2}\right)$$

should be very nearly an integer. In the next chapter we give bounds on the coefficients $c(n)$, and after that we will be in a position to make all this more precise.

We conclude this chapter with some brief remarks on some considerable extensions of Theorems 27 and 29. Firstly one has

Theorem 43 (First Main Theorem of Complex Multiplication) *Let E be an elliptic curve with* $\mathrm{End}(E) \cong \mathcal{O}_K$*, where K is an imaginary quadratic extension of $\mathbb{Q}$. Then*

(i) $[\mathbb{Q}(j(E)) : \mathbb{Q}] = [K(j(E)) : K] = h_K$ *exactly;*

(ii) *If $E_1, \ldots E_h$ is a complete set of representatives for the isomorphism classes of curves with* $\mathrm{End}(E) \cong \mathcal{O}_K$*, then $j(E_1), \ldots, j(E_n)$ is a complete set of* $\mathrm{Gal}(\overline{K}/K)$*-conjugates for $j(E)$;*

(iii) *$K(j(E))$ is the Hilbert Class Field for K.*

For the proof, see for example Silverman [15]. Items (i) and (ii) can be proved essentially by considering the modular equation in more detail- See McKean and Moll [8] for this. It would take us to far afield to even *explain* Item (iii) properly. *Roughly speaking*, the Hilbert Class Field describes the Abelian extensions of K which are also "unramified", these being in 1-1 correspondence with the Galois Group $\mathrm{Gal}(K(j(E))/K)$. This group is naturally isomorphic to the ideal class group $\mathcal{C}(K)$; hence the name. The goal of describing Abelian extensions of imaginary quadratic fields K was known in the 19^{th} century as "Kronecker's Jugendtraum" (Jugendtraum = Youthful Dream). Kronecker had made conjectures in 1860 concerning the use of the j-function in generating such extensions. Theorem 43 was essentially obtained by Weber in 1891. It should be noted that a more general Theorem is possible in which $\mathcal{O}_K$ is replaced by an arbitrary order $\mathcal{O}$ in an imaginary quadratic field.

Secondly one can say even more about $j(E)$ than Theorem 29 does.

Theorem 44 *Let E be an elliptic curve with $\mathrm{End}(E) \cong \mathcal{O}_K$. If $3 \nmid \mathrm{disc}(K)$, then $j(E)$ is a perfect cube.*

For the proof see Cox [5]. There are also some results due to Gross, Zagier and Deuring showing that in some sense only "small" primes may divide $j(E)$ or, more generally, differences $j(E_1) - j(E_2)$. Cox [5] is also a good reference for these.

Chapter 3

In this, the last of our three chapters, we consider the q-expansions of various functions. In particular we will provide a proof of Theorem 37, and will give a precise estimate of how close $e^{\pi\sqrt{163}}$ is to an integer. All of this is most naturally discussed in the language of modular forms, a theory so attractive that I feel no guilt in indulging in a fairly detailed discussion.

3.1 Modular Forms

As we remarked earlier, modular forms of weight $2k$ arise from differential k-forms on the modular curve $X(1) = \mathbb{H}^*/\Gamma(1)$. They are therefore natural objects of study. However we will not set up this correspondence here, choosing instead to define modular forms from scratch.

Definition 45 (Modular Functions) *Let $f : \mathbb{H} \to \mathbb{C}$ be meromorphic and let $k \geq 0$ be an integer. Then f is said to be $\Gamma(1)$-modular of weight $2k$ if it satisfies the transformation law*

$$f(M\tau) = (c\tau + d)^{2k} f(\tau) \tag{3.1}$$

for all $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1)$ and all $\tau \in \mathbb{H}$, and if f is meromorphic at ∞.

We note

(i) The transformation law (3.1) guarantees that $f(\tau + 1) = f(\tau)$ for all $\tau \in \mathbb{H}$, and so f has a q-expansion in the sense of Definition 30. Hence it makes sense to talk about f being meromorphic at ∞.

(ii) In the case $k = 0$, this coincides with our earlier definition of $\Gamma(1)$-modular functions of weight 0.

(iii) Any modular function of odd weight would have to be identically zero (put $M = -I$). This explains the restriction to even weights.

(iv) If f is a modular function of weight $2k$ and g is a modular function of

weight $2m$, then fg is a modular function of weight $2(k+m)$. If f and g are modular functions of the same weight, then f/g is modular of weight 0.

Now it is routine to check that for *any* $f : \mathbb{H} \to \mathbb{C}$ the set of all $M \in \Gamma(1)$ for which (3.1) holds is a subgroup of $\Gamma(1)$. Hence, by Corollary 18, f satisfies (3.1) for all $M \in \Gamma(1)$ if and only if it satisfies it for $M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. To spell it out, f is $\Gamma(1)$-modular iff

$$f(\tau+1) = f(\tau),\ f\left(-\frac{1}{\tau}\right) = \tau^{2k} f(\tau) \tag{3.2}$$

for all $\tau \in \mathbb{H}$ and if f is meromorphic at ∞.

Definition 46 (Modular Forms) *A $\Gamma(1)$-modular function $f : \mathbb{H} \to \mathbb{C}$ which is holomorphic on $\mathbb{H}$ and at ∞ is deemed to be a* modular form. *The set of such f of weight $2k$ forms a $\mathbb{C}$-vector space which we denote $\mathcal{M}_{2k}$. If $f(\infty) = 0$ then f is called a* cusp form *of weight $2k$; these also form a vector space over $\mathbb{C}$, which we denote $\mathcal{M}^0_{2k}$.*

Of course, one can also define modular forms for the subgroups $\Gamma_0(n)$, $\Gamma_1(n)$ and $\Gamma(n)$ of $\Gamma(1)$ discussed earlier.

Examples. (i) The classical examples of modular forms are the Eisenstein Series G_{2k}. For a lattice Λ, define $G_{2k}(\Lambda) = \sum_{\omega \in \Lambda^*} \omega^{-2k}$ (so, with the usual notation, $g_2(\Lambda) = 60G_4(\Lambda)$ and $g_3(\Lambda) = 140G_6(\Lambda)$). For $\tau \in \mathbb{H}$ define

$$G_{2k}(\tau) = G_{2k}(\Lambda_\tau) = \sum_{\substack{m,n \in \mathbb{Z} \\ (m,n) \neq (0,0)}} (m\tau + n)^{-2k}.$$

Now for any $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1)$ the lattices $\Lambda_{M\tau}$ and Λ_τ are homothetic, and an easy calculation shows that in fact $\Lambda_{M\tau} = \frac{1}{c\tau+d}\Lambda_\tau$. But from the form of $G_{2k}(\Lambda)$ it is clear that $G_{2k}(\lambda\Lambda) = \lambda^{-2k} G_{2k}(\Lambda)$ for any $\lambda \in \mathbb{C}^*$ and any lattice Λ. Hence for all $\tau \in \mathbb{H}$ we have

$$G_{2k}(M\tau) = G_{2k}(\Lambda_{M\tau}) = (c\tau+d)^{2k} G_{2k}(\Lambda_\tau) = (c\tau+d)^{2k} G_{2k}(\tau)$$

However we can show that $\lim_{\Im(\tau)\to\infty} G_{2k}(\tau) = 2\zeta(2k)$ using the same method that dealt with g_2 and g_3 in Theorem 21, and also that G_{2k} has no poles in $\mathbb{H}$. Hence G_{2k} is a modular form of weight $2k$, and $G_{2k}(\infty) = 2\zeta(2k)$.

(ii) From the above and note (iii) following Theorem 45, g_2^3 and g_3^2 are both modular forms of weight 12. Hence so is $\Delta = g_2^3 - 27g_3^2$. But we showed that $\lim_{\Im(\tau)\to\infty} \Delta(\tau) = 0$; therefore Δ is actually a cusp form of weight 12.

An obvious question is "How big are $\mathcal{M}_{2k}$ and $\mathcal{M}^0_{2k}$?" We discuss this next.

Theorem 47 *A basis for* $\mathcal{M}_{2k}\,(k \geq 1)$ *is* $\mathcal{B}_{2k} = \{G_4^a G_6^b \,|\, a, b \in \mathbb{N}_0, 2a+3b = k\}$.

Proof We start with a Lemma containing a few properties of G_4 and G_6.

Lemma 48 *G_4 has a simple zero at ρ $(= e^{2\pi i/3})$ and no other zeros in $\mathbb{H}^*/\Gamma(1)$. G_6 has a simple zero at i and no other zeros in $\mathbb{H}^*/\Gamma(1)$.*

Proof The key observations are that $Si = i$ and $ST\rho = \rho$, where $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ as always. Using the modular transformation properties of G_4 and G_6 this gives $G_4(\rho) = (\rho+1)^4 G_4(ST\rho) = \rho^2 G_4(\rho)$ and $G_6(i) = i^6 G_6(Si) = -G_6(i)$. Hence certainly $G_4(\rho) = G_6(i) = 0$. Now we can recast the definition of j in terms of G_4 and G_6 to get

$$j(\tau) = 1728 \frac{20 G_4^3(\tau)}{20 G_4^3(\tau) - 49 G_6^2(\tau)}$$

Recalling that $\Delta = 10800(20G_4^3 - 49G_6^2) \neq 0$ we have $j(\tau) = 0 \Leftrightarrow G_4(\tau) = 0$ and $j(\tau) = 1728 \Leftrightarrow G_6(\tau) = 0$. Since $j : \mathbb{H}^*/\Gamma(1) \to \mathbb{C}_\infty$ is a bijection, this means that ρ is the only zero of G_4 and i is the only zero of G_6. Furthermore, recalling that $v_\rho(j - j(\rho)) = 3$, $v_i(j - j(i)) = 2$ (Theorem 20), it is easy to see that both of these zeros are simple.//

Continuing with the proof of Theorem 47, let us assume that $f \in \mathcal{M}_{2k}$. Choose $c, d \in \mathbb{Z}$ with $2c + 3d = k$ (we do not require $c, d \geq 0$ here). Then $f/G_4^c G_6^d$ is $\Gamma(1)$-modular of weight 0, and so is a rational function in j. But j itself is a rational function in $x(\tau) = G_6^2(\tau)/G_4^3(\tau)$, and so we may write

$$f(\tau) = \frac{G_4^c(\tau) G_6^d(\tau) P(x)}{Q(x)} \tag{3.3}$$

for polynomials P, Q over $\mathbb{C}$. We may assume that P, Q are coprime and also that neither $P(0)$ or $Q(0)$ is zero. Indeed if either $P(x)$ or $Q(x)$ has a factor of x this may be absorbed into the $G_4^c G_6^d$ term. Suppose $Q(x)$ is not just a constant. Then it has some factor $x - \lambda (\lambda \in \mathbb{C})$. But $j : \mathbb{H}^* \to \mathbb{C}$ is surjective and hence so is $x : \mathbb{H}^* \to \mathbb{C}$, and we can find $\tau_0 \in \mathbb{H}^*$ with $x(\tau_0) = \lambda$, giving $Q(x(\tau_0)) = 0$. However $f \in \mathcal{M}_{2k}$ requires that f be holomorphic on $\mathbb{H}^*$, and so τ_0 must also be a zero of the numerator $G_4^c(\tau) G_6^d(\tau) P(x)$ in (3.3). There are three cases:

(i) $P(x(\tau_0)) = 0$. This cannot occur as we supposed P and Q to have no common factor.
(ii) $G_4(\tau_0) = 0$. Then $x(\tau_0) = \infty \neq \lambda$.
(iii) $G_6(\tau_0) = 0$. then $x(\tau_0) = 0$. But we assumed that $Q(0) \neq 0$.

This contradiction forces Q to be constant. It follows from (3.3) that f can be expressed as a finite sum

$$f(\tau) = \alpha_1 G_4^{a_1}(\tau) G_6^{b_1}(\tau) + \ldots + \alpha_n G_4^{a_n}(\tau) G_6^{b_n}(\tau) \tag{3.4}$$

in which $2a_i + 3b_i = k$ for $i = 1, 2, \ldots n$. We may order this so that $a_1 > a_2 > \cdots > a_n$ and $b_1 < b_2 < \cdots < b_n$. If $b_1 < 0$ then f has a pole of order b_1 at i by Lemma 48. But f is holomorphic on $\mathbb{H}^*$, and so this cannot be the case. Therefore $b_1 \geq 0$, and similarly $a_n \geq 0$. It follows immediately that $\mathcal{B}_{2k}$ is a spanning set for $\mathcal{M}_{2k}$. To show that $\mathcal{B}_{2k}$ is a basis, then, we must prove that the elements in it are linearly independent. To see this note that by Lemma 48 the values $\{v_\rho(b_t) \mid b_t \in \mathcal{B}_{2k}\}$ are all different. Order the basis elenents so that $v_\rho(b_1) > \cdots > v_\rho(b_s)$, and suppose we have a relation $\lambda_1 b_1 + \ldots \lambda_s b_s = 0$ with $\lambda_s \neq 0$. Then we get a contradiction on considering the identity

$$\frac{\lambda_1 b_1(\tau) + \cdots + \lambda_s b_s(\tau)}{(\tau - \rho)^{v_\rho(b_s)}} = 0$$

as $\tau \to \rho$. Theorem 47 follows.//

Once we have shown that a weight $2k$ $\Gamma(1)$-modular form is just a differential k-form on $X(1)$, it is possible to give a more conceptual demonstration of this result. See, for example, Silverman [15] for details.

Corollary 49 *For $k \geq 1$,*

$$\dim \mathcal{M}_{2k} = \begin{cases} \lfloor \frac{k}{6} \rfloor & k \equiv 1 (mod\, 6) \\ \lfloor \frac{k}{6} \rfloor & k \not\equiv 1 (mod\, 6) \end{cases}$$

and $\dim \mathcal{M}_{2k}^0 = \dim \mathcal{M}_{2k} - 1$.

Proof The dimension of $\mathcal{M}_{2k}$ is just the number of solutions of $2a + 3b = k$ with $a, b \in \mathbb{N}_0$. The above formulæ can be checked by a 6-case induction on k. $\mathcal{M}_{2k}^0$ is the kernel of the evaluation-at-∞ map $ev_\infty : \mathcal{M}_{2k} \to \mathbb{C}$ and so $\dim \mathcal{M}_{2k}^0 \geq \dim \mathcal{M}_{2k} - 1$. However for each k there are modular forms of weight k which are not cusp forms, namely the Eisenstein Series G_{2k}.//

The fact that these dimensions are so small forces some fascinating identities to hold. For example G_8 and G_4^2 are both in $\mathcal{M}_8$, which has dimension 1. Hence $G_8 = \lambda G_4^2$ for some $\lambda \in \mathbb{C}$. Evaluating at ∞ (see Example (i) above) and using the facts that $\zeta(4) = \frac{\pi^4}{90}$, $\zeta(8) = \frac{\pi^8}{9450}$ we get $\lambda = \frac{3}{7}$. Now we are about to show that the q-series for G_4 and G_8 are

$$\begin{array}{rcl} G_4(\tau) & = & \frac{\pi^4}{45} \left\{ 1 + 240 \sum_{n \geq 1} \sigma_3(n) q^n \right\} \\ G_8(\tau) & = & \frac{\pi^8}{4725} \left\{ 1 + 480 \sum_{n \geq 1} \sigma_7(n) q^n \right\} \end{array} \tag{3.5}$$

where $\sigma_r(n) = \sum_{d|n} d^r$. Substituting this into $G_8 = \frac{3}{7}G_4^2$ and equating coefficients of q^n gives the amazing

$$\sigma_7(n) = \sigma_3(n) + 120 \sum_{m=1}^{n-1} \sigma_3(m)\sigma_3(n-m).$$

This is just one of a whole series of related identities- Shimura [12] gives a general formula. We note in passing that identities such as these can also be obtained by substituting the Laurent Expansion of $\wp_\Lambda$ (Proposition 4) into the differential equation (1.3).

3.2 q-series for j

We now embark on the derivation of the q-expansion of j in a form which will enable us to bound the sum in (2.7). To begin with we need the q-expansions of g_2 and g_3. It turns out to be just as easy to derive the q-expansion of G_{2k}.

3.2.1 q-Expansions of Eisenstein Series

Theorem 50 *Let $k \geq 2$. Then*

$$G_{2k}(\tau) = 2\zeta(2k) + 2\frac{(2\pi i)^{2k}}{(2k-1)!} \sum_{n\geq 1} \sigma_{2k-1}(n) q^n$$

Note that the result generalises (3.5). The quickest proof is that in Serre [11]; this we now give. The starting point is the identity

$$\pi \cot \pi\tau = \frac{1}{\tau} + \sum_{m=1}^{\infty} \left(\frac{1}{\tau + m} + \frac{1}{\tau - m} \right) = \sum_{m\in\mathbb{Z}} \frac{1}{m+\tau} \tag{3.6}$$

which is often established in an elementary complex variable course.

In terms of q, we have easily $\pi \cot \pi z = i\pi - 2i\pi \sum_{n=0}^{\infty} q^n$, and so (3.6) becomes

$$\sum_{m\in\mathbb{Z}} \frac{1}{m+\tau} = i\pi - 2i\pi \sum_{n=0}^{\infty} q^n \tag{3.7}$$

We now differentiate this identity $k-1$ times to get

$$\sum_{m\in\mathbb{Z}} \frac{1}{(m+z)^k} = \frac{1}{(k-1)!}(-2i\pi)^k \sum_{n=1}^{\infty} n^{k-1} q^n \tag{3.8}$$

But we have by definition

$$\begin{aligned} G_{2k}(\tau) &= \sum_{\substack{m,n\in\mathbb{Z}\\(m,n)\neq(0,0)}} (m\tau+n)^{-2k} \\ &= 2\zeta(2k) + 2\sum_{n=1}^{\infty}\sum_{m\in\mathbb{Z}}(m\tau+n)^{-2k} \\ &= 2\zeta(2k) + \tfrac{2(-2\pi i)^{2k}}{(2k-1)!}\sum_{d=1}^{\infty}\sum_{a=1}^{\infty} d^{2k-1}q^{ad} \text{ from (3.8)} \\ &= 2\zeta(2k) + \tfrac{2(2\pi i)^{2k}}{(2k-1)!}\sum_{n=1}^{\infty}\sigma_{2k-1}(n)q^n \end{aligned}$$

as required.//

Suffice it to say that the termwise differentiations and the interchange in the order of summation in the above proof *are* valid: however I prefer not to get caught up in these details here and refer the sceptical reader to a reputable analysis text (e.g. Rudin [10]). The general philosophy is that we have uniform convergence when we need it.

It is convenient to define the *normalized Eisenstein Series* $E_{2k} = \frac{1}{2\zeta(2k)}G_{2k}$. It can be shown (see Silverman [15] or Serre [11]) that $\zeta(2k) = \frac{-(2\pi i)^{2k}}{2(2k)!}B_{2k}$ where the Bernoulli numbers B_m are defined by

$$\frac{x}{e^x-1} = \sum_{m=0}^{\infty} B_m \frac{x^m}{m!}.$$

Hence one has the q-expansion

$$E_{2k}(\tau) = 1 - \frac{4k}{B_{2k}}\sum_{n\geq 1}\sigma_{2k-1}(n)q^n. \tag{3.9}$$

Now we have $\zeta(4) = \frac{\pi^4}{90}$, $\zeta(6) = \frac{\pi^6}{945}$, $g_2 = 60G_4$, $g_3 = 140G_6$, $\Delta = g_2^3 - 27g_3^2$ and $j = \frac{g_2^3}{\Delta}$. This enables us to derive the following entirely calculational proposition, whose proof we omit on grounds of taste.

Proposition 51 *The first few terms in the q-expansions of g_2, g_3, Δ and j are as follows.*

$$\begin{aligned} g_2(\tau) &= \tfrac{4\pi^4}{3}\left\{1 + 24q + 2160q^2 + \cdots\right\} \\ g_3(\tau) &= \tfrac{8\pi^6}{27}\left\{1 - 504q - 16632q^2 + \cdots\right\} \\ \Delta(\tau) &= (2\pi)^{12}\left\{q - 24q^2 + 252q^3 + \cdots\right\} \\ j(\tau) &= \tfrac{1}{q} + 744 + 196884q + \cdots \end{aligned}$$

3.2.2 Jacobi's Product Formula for Δ

We now come to the discussion of some amazing results. It is slightly unfortunate that these will appear a little unmotivated, but the most natural

approach to the whole subject would take us rather far afield. For this approach see Cox [5]; however even this can hardly be described as conceptual.

We start with a definition.

Definition 52 *The Dedekind η-function $\eta(\tau)$ is defined by*

$$\eta(\tau) = q^{1/24} \prod_{n \geq 1} (1 - q^n) \; (\tau \in \mathbb{H})$$

where $q = e^{2\pi i \tau}$ and $q^{1/24} = e^{2\pi i \tau/24}$.

It can be shown that this defines a holomorphic function on $\mathbb{H}$. This function has some remarkable properties, many of which follow from the next Proposition.

Proposition 53 (Transformation Properties of η) *For all $\tau \in \mathbb{H}$ we have*

(i) $\eta(\tau + 1) = e^{2\pi i/24} \eta(\tau)$
(ii) $\eta(-1/\tau) = \sqrt{-i\tau}\eta(\tau)$,

where we take the branch of $\sqrt{}$ which is positive on the positive real axis.

Proof We give an exposition of a slight adaptation of a proof of Siegel [13]. (i) is obvious, so we turn our attention to (ii). Furthermore if we can check this for $\tau = iy$ $(y \in \mathbb{R}^+)$ then by the identity principle it will hold for all $\tau \in \mathbb{H}$. The beauty of doing this is that $q = e^{-2\pi y}$ is then real, and so $\eta(-1/\tau)$, $\sqrt{-i\tau}$ and $\eta(\tau)$ are all real and positive. We may therefore take logarithms of (ii), reducing the identity we are required to prove to

$$\log \eta(i/y) - \log \eta(iy) = \frac{1}{2} \log y. \tag{3.10}$$

Now

$$\begin{array}{rcl} \log \eta(iy) & = & -\frac{\pi y}{12} + \sum_{n=1}^{\infty} \log\left(1 - e^{-2\pi n y}\right) \\ & = & -\frac{\pi y}{12} - \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{e^{-2\pi m n y}}{m} \\ & = & -\frac{\pi y}{12} - \sum_{m=1}^{\infty} \frac{1}{m} \sum_{n=1}^{\infty} e^{-2\pi m n y} \\ & = & -\frac{\pi y}{12} + \sum_{m=1}^{\infty} \frac{1}{m} \frac{1}{1 - e^{2\pi m y}}, \end{array}$$

the interchange in the order of summation being valid because $\sum_{m=1}^{\infty} e^{-2\pi m n y}$ converges uniformly over all $m \in [1, \infty)$. Therefore (3.10) can be written in the form

$$\sum_{m=1}^{\infty} \frac{1}{m} \frac{1}{1 - e^{2\pi m y}} - \sum_{m=1}^{\infty} \frac{1}{m} \frac{1}{1 - e^{2\pi m/y}} - \frac{\pi}{12} \left(y - \frac{1}{y} \right) = -\frac{1}{2} \log y. \tag{3.11}$$

Now comes the clever bit. We write the first two terms as

$$\begin{aligned} & \lim_{N\to\infty}\left\{\left(-\frac{1}{2}\sum_{m=1}^{N}\frac{1}{m}+\sum_{m=1}^{N}\frac{1}{m}\frac{1}{1-e^{2\pi my}}\right)\right. \\ & \qquad\left.-\left(-\frac{1}{2}\sum_{m=1}^{N}\frac{1}{m}+\sum_{m=1}^{N}\frac{1}{m}\frac{1}{1-e^{2\pi m/y}}\right)\right\} \\ = & \lim_{N\to\infty}\left\{\sum_{m=1}^{N}\frac{i}{2m}\cot i\pi my-\sum_{m=1}^{N}\frac{i}{2m}\cot\frac{i\pi m}{y}\right\} \qquad (3.12) \\ = & \lim_{N\to\infty}\left\{\sum_{\substack{m=-N\\ m\neq 0}}^{N}\frac{i}{4m}\cot i\pi my-\sum_{\substack{m=-N\\ m\neq 0}}^{N}\frac{i}{4m}\cot i\pi my\right\} \end{aligned}$$

Now we recall from elementary complex analysis that in summing a series $\sum_{n=-\infty}^{\infty} f(n)$, where $f:\mathbb{C}\to\mathbb{C}$ is analytic except for finitely many poles, it is useful to consider the function $F(z)=f(z)\pi\cot\pi z$. For suitable f this has a pole at $z=n(n\in\mathbb{Z})$ with residue $f(n)$, and so we can get useful information on $\sum_{n=-N}^{N} f(n)$ by considering the integral of $F(z)$ around a suitable square. Let us look at one of the series in (3.12) in this light, say $\displaystyle\sum_{\substack{m=-N\\ m\neq 0}}^{N}\frac{i}{4m}\cot\frac{i\pi m}{y}$.

Let $f(z)=-\frac{i}{4z}\cot\frac{i\pi z}{y}$, so that $F(z)=-\frac{\pi i}{4z}\cot\frac{i\pi z}{y}\cot\pi z$. Here, unlike the usual examples one meets, f has infinitely many poles (at $z=\frac{ky}{i}$ for $k\in\mathbb{Z}$). However the residue of F at $\frac{ky}{i}$ $(k\neq 0)$ can be calculated to be exactly $\frac{i}{4k}\cot i\pi ky$, and so the sum of all these "extra" residues inside a suitable contour C_N looks like giving us exactly the first series

$$\sum_{\substack{m=-N\\ m\neq 0}}^{N}\frac{i}{4m}\cot i\pi my$$

of (3.12)!

Now for our contour C_N we do not in fact take a square with side length $O(n)$ as is usual. Instead we fix C_N to be a square Γ with sides Γ_1, Γ_2, Γ_3 and Γ_4 parallel to the co-ordinate axes and passing through the four points 1, iy, -1, $-iy$. See Figure 3.1. We then "compress" F by an N-dependant factor so that the necessary poles lie inside Γ. this will clearly give the same results, but simplifies discussion of the limit

$$\lim_{N\to\infty}\int_{C_N}F(z)dz.$$

Let us, then, define

$$F_N(z)=-\frac{\pi i}{4z}\cot\frac{i\pi nz}{y}\cot\pi nz$$

*****SEEEBELOWW****.

Figure 3.1: The Contour Γ

where $n = N + \frac{1}{2}$. Inside Γ, F_N has simple poles at $z = \frac{k}{n}$ and at $z = \frac{ky}{in}$ for $k = \pm 1, \ldots \pm n$. The residues at these poles are $-\frac{i}{4k} \cot \frac{\pi i k}{y}$ and $\frac{i}{4k} \cot i\pi k y$ respectively. There is also a triple pole at $z = 0$ with residue $-\frac{\pi}{12}\left(y - \frac{1}{y}\right)$, as one can easily check using the expansion $\cot z = \frac{1}{z} - \frac{z}{3} + O(z^2)$ about 0.

It follows immediately from the Residue Theorem that

$$\frac{1}{2\pi i}\int_\Gamma F_N(z)dz = \sum_{\substack{m=-N \\ m\neq 0}}^{N} \frac{i}{4m} \cot i\pi m y - \sum_{\substack{m=-N \\ m\neq 0}}^{N} \frac{i}{4m} \cot \frac{i\pi m}{y} - \frac{\pi}{12}\left(y - \frac{1}{y}\right).$$

Hence from (3.12) it suffices to show that

$$\frac{1}{2\pi i} \lim_{N\to\infty} \int_\Gamma F_N(z)dz = -\frac{1}{2}\log y \tag{3.13}$$

To prove this we need two Lemmata.

Lemma 54 *F_N is uniformly bounded on Γ independantly of $N \in \mathbb{N}$.*

Proof It clearly suffices to show that both $\cot \pi n z$ and $\cot \frac{i\pi n z}{y}$ are uniformly bounded on Γ. The calculation for $\cot \pi n z$ illustrates the method.

On Γ_1 and Γ_3 we have

$$|\cot \pi n z| = |\cot \pi n(\pm 1 + it)| = |\tan \pi n i t| = \left|\frac{e^{2\pi n t} - 1}{e^{2\pi n t} + 1}\right| < 1$$

On Γ_2 and Γ_4 one has

$$\begin{aligned} |\cot \pi n z| \;=\; |\cot \pi n(t \pm iy)| &= \left|\tfrac{1+e^{\pm 2y\pi n}e^{-2it\pi n}}{1-e^{\pm 2y\pi n}e^{-2it\pi n}}\right| \\ &\leq \tfrac{1+e^{\pm 2y\pi n}}{|1-e^{\pm 2y\pi n}|} \\ &\leq \max\left\{\tfrac{1+e^{\pm y}}{|1-e^{\pm y}|}, 1\right\}, \end{aligned}$$

this last step following from some obvious properties of the graph of $\frac{1+x}{|1-x|}$ on $x > 0$.//

Lemma 55 *As $N \to \infty$, $F_N(z)$ converges pointwise on Γ to the function G defined on Γ as follows:*

$$G(z) = \begin{cases} -\frac{\pi i}{4z} & (\Re z \cdot \Im z > 0) \\ \frac{\pi i}{4z} & (\Re z \cdot \Im z < 0) \\ 0 & (\Re z \cdot \Im z = 0) \end{cases}$$

Proof Just check the pointwise limits of $\cot \pi n z$ and $\cot \frac{i\pi n z}{y}$ separately using $\cot(\theta) = \frac{1+e^{-2i\theta}}{1-e^{-2i\theta}}$.//

We can now use the Bounded Convergence Theorem of Lebesgue Integration Theory and a short calculation to deduce that

$$\frac{1}{2\pi i} \lim_{N\to\infty} \int_\Gamma F_N(z)dz = \lim_{N\to\infty} \int_\Gamma G(z)dz = -\frac{1}{2} \log y$$

Hence (3.13) holds and we have proved Proposition 53.//

An immediate consequence is Jacobi's Famous Product Expansion for $\Delta(\tau)$.

Theorem 56 (Jacobi's Product Expansion for $\Delta(\tau)$) *For all $\tau \in \mathbb{H}$,*

$$\Delta(\tau) = (2\pi)^{12}\eta(\tau)^{24} = (2\pi)^{12} q \prod_{n\geq 1} (1-q^n)^{24}.$$

Proof In view of Proposition 53 it is trivial to verify that $\eta(\tau)^{24}$ satisfies (3.2) with $k = 6$. Furthermore directly from Definition 52 we see that $\lim_{\Im(\tau)\to\infty} \eta(\tau) = 0$. It follows that $\eta(\tau)^{24}$ is a cusp form of weight 12. But from Corollary 49 we see that $\dim_{\mathbb{C}} \mathcal{M}_{12}^0 = 1$. Hence we must have $\Delta(\tau) = \lambda\eta(\tau)^{24}$ for some $\lambda \in \mathbb{C}$, and it is immediate from the q-expansions that $\lambda = (2\pi)^{12}$.//

We note that the Dedekind η-function can also be used to construct modular forms for congruence subgroups of $\Gamma(1)$; for example $\eta(\tau)^2\eta(11\tau)^2$ is a weight 2 cusp form for $\Gamma_0(11)$. We can also find expressions for j in terms of η- an example is

$$j(\tau) = \frac{\eta(\frac{1}{2}\tau)^{16}}{\eta(\tau)^{16}} + \frac{16\eta(\tau)^8}{\eta(\tau/2)^8}$$

which may be *verified* by checking that the right-hand side is $\Gamma(1)$-invariant and than inspecting the first few terms of the q-expansion. For a more perspicacious derivation, see Cox [5].

It follows from Theorem 56 that $\Delta(\tau) = (2\pi)^{12} \sum_{n\geq 1} \boldsymbol{\tau}(n) q^n$ for integers

$\boldsymbol{\tau}(n)$. $\boldsymbol{\tau}$ is called the Ramanujan Function.

We can now prove (at last) Theorem 37. Note, however, that this could have been done, albeit a little more lengthily than we are about to, directly from the q-expansions of g_2 and g_3.

Proof of Theorem 29 We compute using the q-expansion of g_2 and Theorem 56

$$j(\tau) = \frac{1728 g_2^3(\tau)}{\Delta(\tau)} = \frac{\left\{1 + 240 \sum_{n\geq 1} \sigma_3(n) q^n\right\}^3}{q \prod_{n\geq 1}(1-q^n)^{24}} \tag{3.14}$$

which makes the whole of Theorem 37 clear.//

3.2.3 Properties of $c(n)$ and $\boldsymbol{\tau}(n)$

Now is perhaps the time to mention some remarkable facts about the coefficients $c(n)$ and $\boldsymbol{\tau}(n)$. The following is just a sample:

Identities for $\boldsymbol{\tau}(n)$

(1) For all $m, n \in \mathbb{Z}$ with $\gcd(m,n) = 1$ we have $\boldsymbol{\tau}(mn) = \boldsymbol{\tau}(m)\boldsymbol{\tau}(n)$.

(2) For all primes p and all $e \geq 1$ we have $\boldsymbol{\tau}(p^e)\boldsymbol{\tau}(p) = \boldsymbol{\tau}(p^{e+1}) + p^{11}\boldsymbol{\tau}(p^{e-1})$. These were conjectured by Ramanujan and proved by Mordell. See Silverman [15].

Congruences for $\boldsymbol{\tau}(n)$

(3) For all $n \in \mathbb{N}$, we have $\boldsymbol{\tau}(n) \equiv \sigma_{11}(n) \pmod{691}$.

Proof Using $\dim_{\mathbb{C}} \mathcal{M}_{12} = 2$ and the values $B_4 = -\frac{1}{30}$, $B_6 = -\frac{1}{42}$, $B_{12} = -\frac{691}{2730}$ in (3.9) shows after a little work that $691E_6 = 250E_3^2 + 441E_2^3$. One also has

$$\sum \boldsymbol{\tau}(n) q^n = (2\pi)^{-12}\Delta = g_2^3 - 27 g_3^2 = \frac{E_2^3 - E_3^2}{1728}.$$

Eliminating gives

$$\begin{aligned} 2^6 3^5 7^2 \textstyle\sum_{n\geq 1} \boldsymbol{\tau}(n) q^n &= 691(E_6 - E_3^2) \\ &\equiv 65520 \textstyle\sum_{n\geq 1} \sigma_{11}(n) q^n \pmod{691} \end{aligned}$$

with an obvious notation, on noting that $E_3 = 1 - 504 \sum \sigma_5(m) q^m$ has integer coefficients, and that $E_6 = 1 + \frac{65520}{691} \sum \sigma_{11}(m) q^m$. The result follows on equating coefficients and noting that $2^6 3^5 7^2 \equiv 65520 \not\equiv 0 \pmod{691}$.//

Inequality for $\tau(n)$

(4) $|\tau(p)| \leq 2p^{11/2}$ for all primes p. This was shown by Deligne as a consequence of his proof of the "Riemann Hypothesis" for varieties over finite fields.

Congruences for $c(n)$

(5) If $p = 2, 3, 5, 7, 11$ then $n \equiv 0 \pmod{p^e} \Rightarrow c(n) \equiv 0 \pmod{p^e}$

Monstrous Moonshine

(6) There is a surprising connection between the $c(n)$ and the dimensions of the irreducible characters of the Monster Simple group. This was first discussed in Conway-Norton [4].

3.3 Bounds on the coefficients $c(n)$

Formula 3.14 enbles us to bound the coefficients $c(n)$ accurately enough for us to establish the sum in (2.7) to our satisfaction. First of all we look at the expansion

$$\prod_{n\geq 1}(1-q^n)^{-24} = \sum_{n=1}^{\infty}\gamma(n)q^n.$$

We have the following estimate for the $\gamma(n)$.

Theorem 57 *For all $n \geq 1$, $\gamma(n) < e^{4\pi\sqrt{n}}$.*

Proof This is proved in a very similar manner to the proof in Apostol [1] that $p(n) < e^{\pi\sqrt{2n/3}}$, where p is the partition function.

Let $F(q) = \prod_{n\geq 1}(1-q^n)^{-24}$. It can be shown that this converges on $|q| < 1$. Then

$$\begin{aligned} \log F(q) &= -24\sum_{n\geq 1}\log(1-q^n) \\ &= 24\sum_{n\geq 1}\sum_{m\geq 1}\frac{q^{mn}}{m} \\ &= 24\sum_{m\geq 1}\frac{1}{m}\cdot\frac{q^m}{1-q^m} \end{aligned} \tag{3.15}$$

the interchange in the order of summation being valid since the series

$$\sum_{m\geq 1}\frac{q^{mn}}{m}$$

converges uniformly on any compact subset of the open unit disc. Now if $0 < q < 1$ we have

$$mq^{m-1}(1-q) < (1+q+\ldots+q^{m-1})(1-q) = 1-q^m$$

and so

$$\frac{1}{m}\left(\frac{q^m}{1-q^m}\right) < \frac{q}{m^2(1-q)}.$$

Hence from (3.15) one has

$$\log F(q) < 4\pi^2\left(\frac{q}{1-q}\right). \tag{3.16}$$

Now each $\gamma(n)$ is non-negative, and so

$$\gamma(n)q^n < \sum_{n=1}^{\infty}\gamma(n)q^n = F(q).$$

Hence

$$\begin{aligned}\log\gamma(n) &\leq \log F(q) - n\log q \\ &< 4\pi^2\left(\tfrac{q}{1-q}\right) + n\left(\tfrac{1-q}{q}\right),\end{aligned} \tag{3.17}$$

where we have used the inequality $\log q < \frac{1-q}{q}$, valid for $q \in (0,1)$. Note that (3.17) is valid for *any* $q \in (0,1)$. The tightest bound results from putting $q = \frac{\sqrt{n}}{2\pi+\sqrt{n}}$, giving $\log\gamma(n) < 4\pi\sqrt{n}$.//

Our next job is to (very crudely) estimate the coefficients $\theta(n)$ in the expansion

$$\left\{1 + 240\sum_{n\geq 1}\sigma_3(n)q^n\right\}^3 = \sum_{n=1}^{\infty}\theta(n)q^n.$$

First of all observe that

$$\sigma_3(n) \leq 1^3 + 2^3 + \ldots + n^3 = \frac{1}{4}n^2(n+1)^2. \tag{3.18}$$

Hence

$$\begin{aligned}\theta(n) &< 240^3 \sum_{\substack{l+k+m=n \\ 0\leq l,k,m\leq n}} \sigma_3(l)\sigma_3(k)\sigma_3(m) \\ &\ll \tfrac{240^3}{4^3}n^6(n+1)^6 \cdot \#\left\{l,k,m \in \mathbb{Z} \mid 0 \leq l,k,m \leq n, l+k+m=n\right\} \\ &= 60^3 n^6(n+1)^6 \left(\begin{array}{c} n+2 \\ 2 \end{array}\right),\end{aligned} \tag{3.19}$$

where we have set $\sigma_3(0) = 1$.

Finally, using Theorem 57, (3.14) and the above we have

$$\begin{aligned}c(n) &= \sum_{\substack{k+m=n+1 \\ k,m\geq 0}} \gamma(k)\theta(m) \\ &\leq (n+2)\max_{k\leq n+1}\gamma(k)\max_{m\leq n+1}\theta(m) \\ &\leq 60^3(n+1)^6(n+2)^6\left(\begin{array}{c} n+3 \\ 2 \end{array}\right)e^{4\pi\sqrt{n+1}}.\end{aligned} \tag{3.20}$$

This does not compare too unfavourably with the asymptotic formula of Petersson [9] for $c(n)$, viz

$$c(n) \sim \frac{e^{4\pi\sqrt{n}}}{\sqrt{2}n^{3/4}} \text{ as } n \to \infty$$

Let us now look at the implications of (3.20) for the sum S in (2.7). We have

$$\begin{aligned} |S| &= \left|\sum_{n=1}^{\infty} c(n)(-1)^n e^{-n\pi\sqrt{163}}\right| \\ &\leq \left|196884e^{-\pi\sqrt{163}} - c(2)e^{-2\pi\sqrt{163}}\right| + \sum_{n=3}^{\infty} S_n, \end{aligned} \tag{3.21}$$

where $S_n = 60^3(n+1)^6(n+2)^6 \begin{pmatrix} n+3 \\ 2 \end{pmatrix} e^{4\pi\sqrt{n+1}-n\pi\sqrt{163}}$. Now by suitable means (a pocket calculator, or some clever work with inequalities) one can check that $S_2 < 3 \times 10^{-13}$, so

$$c(2)e^{-2\pi\sqrt{163}} < S_2 < 3 \times 10^{-13}. \tag{3.22}$$

Furthermore for $n \geq 2$ one has

$$\begin{aligned} \tfrac{S_{n+1}}{S_n} &= \tfrac{(n+3)^6(n+4)}{(n+1)^6(n+2)} \cdot e^{4\pi\{\sqrt{n+2}-\sqrt{n+1}\}-\pi\sqrt{163}} \\ &\leq \tfrac{5^6 \cdot 6}{3^6 \cdot 4} \cdot e^{4\pi(2-\sqrt{3})-\pi\sqrt{163}} \\ &\approx 1.1 \times 10^{-16} \\ &\ll \tfrac{1}{4}. \end{aligned} \tag{3.23}$$

Now finally note that

$$7 \times 10^{-13} < 196884e^{-\pi\sqrt{163}} < 8 \times 10^{-13}$$

and so we have from (3.21), (3.22) and (3.23)

$$|S| \leq 8 \times 10^{-13} + 10^{-13} < 10^{-12}. \tag{3.24}$$

Recalling the equation (2.7), this implies

Theorem 58 $j\left(\frac{1+\sqrt{-163}}{2}\right) = [\![-e^{\pi\sqrt{163}} + 744]\!] = -640320^3$
$= -2^{18}5^3 3^3 23^3 29^3$.

Note that we have verified Theorem 44 in this case.

Theorem 59 *$e^{\pi\sqrt{163}}$ is within 10^{-12} of an integer.//*

Bibliography

[1] Apostol, T.M. *Introduction to Analytic Number Theory*, Springer-Verlag, 1976.

[2] Apostol, T.M. *Modular Functions and Dirichlet Series in Number Theory*, GTM41, Springer-Verlag, New York 1976.

[3] Cohn, H. *Introduction to the construction of Class Fields*, Dover 1994.

[4] Conway, J.H. and Norton, S.P. *Monstrous Moonshine*, Bull. Lond. Math. Soc. 11 (1979), pp308 - 339.

[5] Cox, D.A. *Primes of the form* $x^2 + ny^2$, Wiley and Sons 1989.

[6] Hüsemoller, D *Elliptic Curves*, GTM111, Springer Verlag, New York 1986.

[7] Jones, G.A. and Singerman, D. *Complex Functions*, CUP 1987.

[8] McKean, H.P. and Moll, V. *Elliptic Curves*, CUP 1997.

[9] Petersson, H. *Über die Entwicklungskoeffizienten der automorphen Formen*, Acta. Math. 58 (1932) 169-215.

[10] Rudin, W. *Principles of Mathematical Analysis*, McGrawHill 1976.

[11] Serre, J-P. *A course in Arithmetic*, GTM7, Springer-Verlag 1973.

[12] Shimura, G. *Introduction to the Arithmetic Theory of Automorphic Forms*, Princeton University Press, Princeton NJ, 1971.

[13] Siegel, C.L. *A simple proof that* $\eta(-1/\tau) = \eta(\tau)\sqrt{\tau/i}$, Mathematika 1 (1954) 4.

[14] Silverman, J.H. *The Arithmetic of Elliptic Curves*, GTM106, Springer-Verlag 1986.

[15] Silverman, J.H. *Advanced Topics in the Arithmetic of Elliptic Curves*, GTM151, Springer-Verlag 1994.