On the *abc*-conjecture

by: Keldon W. Drudge

Department of Mathematics and Statistics,
McGill University, Montreal

July 1995

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Master of Science

Abstract

This thesis examines the *abc*-conjecture, a conjectured diophantine inequality which makes a connection between the operations of addition and multiplication (specifically, prime factorization) in number fields. After examining the context and motivation of the conjecture, we go on to prove two very different implications of it - one algebraic and one analytic.

Résumé

Cette thèse examine la conjecture *abc*, une inégalité diophantine conjecturé qui fait un rapport entre les operations d'addition et de multiplication (spécifiquement, decomposition en facteurs premiere) dans l'anneau des nombres entier et les corps de nombres. D'abord, nous examinons le contexte et motivation de la conjecture, et puis nous prouvons deux imlications tres differente d'elle - une algebraique et une analytique.

Acknowledgements

I would like to thank Bill Boshuck and Hassan Daghigh for taking the time to discuss various concepts and ideas (both related to this thesis and not) with me, Bill Boshuck and David Alexander for providing me with advice regarding the finer points of AMS-LaTeX, and of course my supervisor Ram Murty for giving me the opportunity to study these topics, and for providing me with much useful information along the way.

Contents

Chapter 1

Introduction

1.1 Motivation and Statement of Conjecture

The *abc*-conjecture is one of the most interesting recent conjectures in number theory. The past decade or so has been marked by great progress in number theory, and concurrent statements of conjectures the resolution of which would seem to require further steps forward. The *abc*-conjecture sits among this group of statements in a web of equivalences and implications which give, if nothing else, at least heuristic evidence for their truth. The goal of this thesis is to highlight some of the many implications of this conjecture; in the process much interesting mathematics is discussed. We begin by stating the conjecture. Like many other famous conjectures in number theory, it appears innocuous.

> **(*abc*-Conjecture)** For every $\epsilon > 0$ there exists a constant $M = M_\epsilon$ such that if a, b, and c are coprime positive integers satisfying $a + b = c$, then
>
> $$c \leq M(\prod_{p|abc} p)^{1+\epsilon}.$$

For convenience, we will often denote the product on the right hand side of the inequality by $G(a, b, c)$, and we will use the usual notation $\ll$ or $\ll_\epsilon$ to mean 'less than or equal to up to a constant depending on ϵ'.

This statement was first given in the above form in D.W. Masser's 1985 paper [21]; Masser refined a question of Oesterlé in analogy with a result of R.C. Mason, which in [20] is used to find constructive algorithms for finding solutions to polynomial equations such as $f(x, y) = 1$ over function fields. By way of motivation, we present here a simple version of this inequality, as well as its (elementary) proof.

Theorem 1 (Mason) *Let $a(z)$, $b(z)$, and $c(z)$ be non-constant relatively prime polynomials in $\mathbb{C}[z]$. If $a + b + c = 0$ then*

$$\max\{\deg(a), \deg(b), \deg(c)\} < r,$$

where r is the number of distinct roots of abc.

Proof: Let a, b c and r be as in the statement of the theorem, and assume without loss of generality that $\deg(c) \geq \deg(a)$ and $\deg(c) \geq \deg(b)$. We will denote the (formal) derivative of a polynomial $f \in \mathbb{C}[z]$ by f'. Since a and b have no common factors and $\deg(a') < \deg(a)$, $\deg(b') < \deg(b)$,

$$\frac{a}{b} \neq \frac{a'}{b'},$$

or in other words $ab' - ba' \neq 0$. If for some $\alpha \in \mathbb{C}$, $(z-\alpha)^e | a$, $(z-\alpha)^{e-1} | (ab' - ba')$ and therefore

$$a(z) \mid (ab' - ba') \prod_{a(\alpha)=0} (z - \alpha).$$

The same reasoning holds for b, and furthermore, since $a + b + c = 0$

$$ab' - ba' = a(-a' - c') - (-a - c)a' = ca' - ac'$$

so the same reasoning also holds for c and since a, b and c have no common factors we end up with

$$a(z)b(z)c(z) \mid (ab' - ba') \prod_{\substack{a(\alpha)=0 \\ b(\alpha)=0 \\ c(\alpha)=0}} (z - \alpha),$$

so $\deg(a) + \deg(b) + \deg(c) \leq \deg(a) + \deg(b) - 1 + r$. Subtracting $\deg(a) + \deg(b)$ from both sides gives the result.

Notice that this theorem provides a quick solution to Fermat's last theorem in the function field case - that is, applying theorem 1 to the equation $x^n + y^n = z^n$ with x, y and z coprime non-constant polynomials in $\mathbb{C}[z]$ shows immediately that $n < 3$. This gives a hint as to the power of the above inequality, and indeed, Mason's book [20] consists largely of applications of the inequality to the study of diophantine equations over function fields. This in itself provides motivation for the proof of a similar inequality in the number field case.

The *abc*-conjecture arises when one attempts to transplant theorem 1 to number fields, or specifically to the rational integers. The analog of the factors $(z - \alpha)$ are the prime numbers, so we have the following 'literal' translation of theorem 1, in which we denote by $\nu(n)$ the number of prime factors of n, counted with multiplicity:

> (A) *If a, b, and c are relatively prime integers whose sum is zero, then* $\max\{\nu(a), \nu(b), \nu(c)\} < r(abc)$,

where $r(n)$ is the number of primes dividing n. This statement is manifestly false in $\mathbb{Z}$, however, as can be seen by the following

Lemma 1 *If $x \equiv y$ (mod p) then $x^{p^n} \equiv y^{p^n}$ (mod p^{n+1}).*

Proof: We proceed by induction on n, the case $n = 0$ being the hypothesis. If $x^{p^{n-1}} \equiv y^{p^{n-1}}$ (*mod* p^n) then since

$$x^{p^n} - y^{p^n} = (x^{p^{n-1}} - y^{p^{n-1}})(x^{(p-1)p^{n-1}} + x^{(p-2)p^{n-1}}y^{p^{n-1}} + \cdots + y^{(p-1)p^{n-1}})$$

and the sum on the right hand side has p terms, all congruent modulo p, the lemma is proved.

This lemma provides counterexamples to (A) by examining the equation $x^{p^n} = y^{p^n} + dp^{n+1}$; for simplicity put $x = 5$, $y = 2$, $p = 3$ (we could of course use $x = 4$, $y = 1$ and $p = 3$, but since 1 is a unit in $\mathbb{Z}$ and the inequality proven by Mason disallows units in $\mathbb{C}[z]$, this may be regarded as 'cheating'). Then (A) implies that $3^n < 3 + r(d)$, but $d < 5^{3^n}$, so d must have at most $o(3^n)$ distinct prime divisors, since

$$\lim_{n \to \infty} \frac{5^n}{2 \times 3 \times \cdots \times p_n} = 0,$$

where p_n is the n^{th} prime. Note that this example also shows that modifying (A) by multiplying the right hand side of the inequality by a constant is not sufficient. Since rational prime numbers have a natural 'height' (the height of p is $\log p$) whereas prime elements of $\mathbb{C}[z]$ (that is, polynomials of the form $z - \alpha$ for $\alpha \in \mathbb{C}$) do not, it seems reasonable to modify (A) by incorporating the size of the primes into the inequality - this modification gives us

> (B) *If a, b, and c are relatively prime integers whose sum is zero, then*
>
> $$\max\{\log|a|, \log|b|, \log|c|\} < \sum_{p|abc} \log p.$$

This attempt is also derailed by the equation $5^{3^n} = 2^{3^n} + d3^n$ however, since when applied to this equation it predicts that

$$3^n \log 5 < \log 30 + \log d < \log 30 + 3^n \log 5 - n \log 3$$

which evidently fails for $n > 3$. However, this attempt is 'close' here, in the sense that multiplying the right hand side of the predicted inequality by any constant greater than one gives a correct prediction - at least for large n. Thus we introduce a multiplicative constant greater than one, and an additive constant to compensate for the cases where n is small, and obtain

> (C) *For every $\epsilon > 0$ there exists a constant M'_ϵ such that if a, b, and c are relatively prime integers whose sum is zero,*
>
> $$\max\{\log|a|, \log|b|, \log|c|\} < M'_\epsilon + (1 + \epsilon) \sum_{p|abc} \log p.$$

Of course, exponentiating both sides of this gives the *abc*-conjecture as stated above. Now, if one were to to begin this process with an inequality closer in generality to the one in Mason [20, Chapter 1, lemma 2] and attempt to translate it to the number field case, one would end up with a conjecture similar to the *abc*-conjecture but for number fields, whose statement is as follows:

> **(*abc*-Conjecture for number fields)** For all number fields k and real numbers $\epsilon > 0$, there exists $M = M_{k,\epsilon} > 0$ such that if $a, b, c \in k$, $a + b + c = 0$, one has
>
> $$H(a,b,c) \leq MG(a,b,c)^{1+\epsilon}.$$

Here,

$$H(a,b,c) = \prod_{v} \max\{\|a\|_v, \|b\|_v, \|c\|_v\}$$

is the *height* of $\{a, b, c\}$ and

$$G(a,b,c) = \prod_{\mathfrak{p} \in S} \|\mathfrak{p}\|_{\mathfrak{p}}$$

is their *conductor*. The product in H runs over all non-Archimedean primes $v = \mathfrak{p}$ with $\| x \|_{\mathfrak{p}} = N_{k/\mathbb{Q}}(\mathfrak{p})^{-1/[k:\mathbb{Q}]}$ and over all Archimedean primes $v = \sigma : k \to \mathbb{C}$ with $\|x\|_\sigma = |\sigma(x)|^{1/[k;\mathbb{Q}]}$, and S is the set of prime ideals $\mathfrak{p}$ for which $\| a \|_{\mathfrak{p}}$, $\| b \|_{\mathfrak{p}}$, $\| c \|_{\mathfrak{p}}$ are not all equal. The *uniform abc*-conjecture for number fields is the slightly stronger statement that there exists some $E > 0$ such that for every $\epsilon > 0$, the constant $M_{k,\epsilon}$ can be taken to be $\Delta_{[k:\mathbb{Q}]}^{E} M_\epsilon$ with $\Delta_{[k:\mathbb{Q}]} = Disc(k/\mathbb{Q})^{1/[k:\mathbb{Q}]}$ (the 'normalized discriminant' of k). In this thesis, unless otherwise stated *abc*-conjecture or simply *abc* will refer to the *abc*-conjecture for the integers.

1.2 Plausibility and Optimality

Given that statements (A) and (B) from the previous section were easily disposed of, is there any reason to believe that counterexamples to the *abc*-conjecture do not also exist? Of course, the presence of the constants and the

ϵ make it much more difficult to disprove, since one would need an infinite sequence of triples $\{x, y, z\}$ such that $\gcd(x, y, z) = 1$, $x + y = z$ and $x > G(a, b, c)^{1+\epsilon}$ for some $\epsilon > 0$. In their paper [29], Stewart and Tijdeman remark that although a proof of *abc* seems hopeless (by virtue of its implications - see section 1.3), it may be possible to disprove it, and they show that the conjecture is close to being 'best possible'. Specifically, they show that for any $\delta > 0$, there exist infinitely many triples of coprime positive integers $\{x, y, z\}$ with $x + y = z$ and

$$x > G \exp\left((4-\delta)\frac{\sqrt{\log G}}{\log\log G}\right) \tag{1.1}$$

where $G = G(x, y, z)$. Of course, this does not disprove the conjecture, since for any $\epsilon, \delta > 0$

$$\exp\left((4-\delta)\frac{\sqrt{\log G}}{\log\log G}\right) = o(G^{\epsilon})$$

as $G \to \infty$. Still, it does provide the following

Proposition 1 *As $\epsilon \to 0$, the constants M_ϵ in the statement of the abc-conjecture tend to infinity.*

Proof: This follows easily from the result of Stewart and Tijdeman. Put $G = G(x, y, z)$. Assume that some M works for every ϵ and pick $\{x, y, z\}$ satisfying (1.1) and large enough so that $\log G > (\log M)^2$. Then pick $\epsilon > 0$ to be less than

$$\frac{1}{\log G}\left(\frac{\sqrt{\log G} - \log M}{\log\log G}\right),$$

and rearranging the inequality from the *abc*-conjecture shows that M fails for this value of ϵ.

Thus, the *abc*-conjecture is at least 'close' to being best possible. The question of plausibility, however, is more subjective. Of course, the fact that the statement has not been disproved for ten years provides some measure of reassurance, but not very much, the integers being infinite. However, *abc* has been proven equivalent to several other conjectures in number theory, and progress toward its resolution has been made by Stewart and Tijdeman in [29] and by Stewart and Yu in [30] (in the latter it is proven that for all $\epsilon > 0$ there exists an effectively computable constant M_ϵ such that for $a, b, c \in \mathbb{Z}$, $a + b = c$ and $\gcd(a, b, c) = 1$ implies $\log c \leq M_\epsilon(G(a, b, c))^{2/3}$). We now give two conjectures equivalent to *abc*, assuming that the reader is familiar with some notions about elliptic curves (see [27] for definitions).

> **(Degree conjecture)** If $E/\mathbb{Q}$ is a modular elliptic curve with modular parameterization $\phi : X_0(N) \to E$ where N is the conductor of E, then
>
> $$\log\left(\frac{\deg \phi}{c_E^2}\right) = O(\log genus(X_0(N))).$$

Here, $X_0(N) = (\Gamma_0(N) \setminus \mathcal{H})^*$ is (the compactification of) the upper half plane modulo the left action of the congruence subgroup $\Gamma_0(N)$, and c_E is a constant associated to the map ϕ (the pull back of the Nerón differential by ϕ to $X_0(N)$ is c_E times a normalized newform of weight two and level N). This was proven equivalent to *abc* assuming the Taniyama-Shimura conjecture that all elliptic curves over $\mathbb{Q}$ are modular, by L. Mai and R. Murty in [19] using a version of the Phragmen-Lindelof theorem and the following conjecture which was proven equivalent to *abc* by G. Frey in [7].

> **(Height Conjecture)** If $E/\mathbb{Q}$ is an elliptic curve with conductor N then $h(E) = O(\log N)$.

Here, $h(E)$ is the Faltings height of the curve E (for definition, see [7] or [19]). Very recently, R. Murty has shown that if a prime p divides the degree of the parameterization ϕ, then $p \leq N^2$. In [31], Vojta proves the equivalences of the height conjecture with *abc* and with two other conjectures, one (Szpiro's conjecture) about the discriminant of elliptic curves, and the other (Hall-Lang-Waldschmidt-Szpiro conjecture) on solutions of the family of diophantine equations of the form $Ax^m + By^n = z \neq 0$ for rational integers A and B. Vojta also shows that these conjectures are all consequences of his 'General Conjecture' which is itself a generalisation of various known results in algebraic geometry and diophantine approximation.

1.3 Easy Consequences

The *abc*-conjecture is very well suited to the study of solutions of binary diophantine equations in the integers, and in this section we present, as examples of the conjecture's versatility in this area, a few propositions which would be immediate were the truth of the conjecture known. The first originally appeared in Marius Overholt's paper [25]; we present here a slightly strengthened version.

Proposition 1 (Overholt) *If the abc-conjecture is true, then for any $k > 1$, the equation $n! + 1 = m^k$ has only finitely many solutions in rational integers.*

Proof: The proof depends on the fact that

$$\prod_{p \le n} p \le 4^n. \tag{1.1}$$

We present a simple proof of this fact due to Erdös. (1.1) is clearly true for $n = 2$, so proceeding by induction assume that it is true for all integers less than n. If n is even, there is nothing to prove, so assume $n = 2m + 1$ is odd. Now it is easily proven (by induction on m) that

$$\binom{2m+1}{m} \le 4^m;$$

and this quantity is divisible by all primes between $m+2$ and $2m+1$ inclusive. Thus

$$\prod_{p \le n} p = \prod_{p \le m+1} p \prod_{m+2 \le p \le n} p \le 4^{m+1} 4^m = 4^n$$

which completes the proof. To prove the proposition, we apply the *abc*-conjecture to the equation $n! + 1 = m^k$, yielding

$$m^k \ll (m \prod_{p \le n} p)^{1+\epsilon} \le (m4^n)^{1+\epsilon} \tag{1.2}$$

for every $\epsilon > 0$. Also, since for $n \ge 3$, $n! \ge 4n^n e^{-n}$ (this is easily proven by induction, using the fact that for every n, $\left(1 + \frac{1}{n}\right)^n < e$), we have that

$m^k \geq n! \geq 4n^n e^{-n}$. Inserting this into (1.2) gives $4n^n e^{-n} m^{-1-\epsilon} \ll 4^{n(1+\epsilon)}$. Using the obvious bound $n! < n^n - 1$ so that $m < n^{n/k}$, we have $n^{n-(1+\epsilon)n/k} \ll 4^{n(1+\epsilon)} e^n$, and on taking logarithms of both sides this becomes

$$n\left(1 - \frac{1+\epsilon}{k}\right)\log n \leq n \log(4^{1+\epsilon} e) + c_0$$

for some constant c_0. This provides a bound on n, and proves the proposition.

We now show that the *abc*-conjecture implies that there exist only finitely many counterexamples to Fermat's last theorem. In fact, we have the following:

Proposition 2 *Given q, r and s positive integers satisfying*

$$\frac{1}{q} + \frac{1}{r} + \frac{1}{s} < 1, \tag{1.3}$$

The abc-conjecture implies that there exist only finitely many triples of integers (x, y, z) satisfying

$$x^q + y^r = z^s. \tag{1.4}$$

Proof: Since $\frac{1}{q} + \frac{1}{r} + \frac{1}{s} < 1$, there exists an $\epsilon > 0$ such that $qrs > (qs + rs + qr)(1 + \epsilon)$. Applying the *abc*-conjecture with this ϵ to the equation (1.4) yields

$$z^s \ll \Big(\prod_{p|xyz} p\Big)^{1+\epsilon} \leq (xyz)^{1+\epsilon} \leq (z^{s/q} z^{s/r} z)^{1+\epsilon},$$

and isolating z shows that

$$z^{\frac{qrs}{(qs+rs+pq)(1+\epsilon)} - 1} \ll 1,$$

so there exist only finitely many solutions for fixed q, r, and s.

Corollary *The abc-conjecture implies that there exist at most finitely many quadruples (x, y, z, n) where x, y, and z are positive integers and $n \geq 4$ is a positive integer and $x^n + y^n = z^n$.*

Proof: The proof of the proposition shows that in any solution $x^n + y^n = z^n$, we must have $z^{-1+n/(3+3\epsilon)} \ll 1$, and since in any solution we must have $z \geq 2$, there exists an N such that for $n > N$, $x^n + y^n = z^n$ has no solutions. For a fixed n *abc* implies that the equation can have only finitely many solutions, so this proves the corollary.

It has been conjectured both by Erdös [6] and Mollin and Walsh [22] that there exist only finitely many triples of consecutive powerful integers, where an integer n is said to be *powerful* if for every prime p, $p|n$ implies $p^2|n$. It was shown by Granville [8] that the *abc*-conjecture implies the truth of this statement; we give a different proof of this implication here.

Proposition 3 *If the abc-conjecture is true, then there exist only finitely many triples of powerful integers.*

Proof: Assume that $l-1, l$ and $l+1$ are all powerful. Then since no powerful integer can be congruent to two modulo four, $l-1, l$ and $l+1$ must be congruent to 3, 0 and 1 respectively modulo 4, so that $l = 4n$ for some integer n. We know that for any n, $\{4n, 4n^2-1, 4n^2+1\}$ give a Pythagorean triple. Apply the *abc*-conjecture to the equation

$$(4n)^2 + (4n^2-1)^2 = (4n^2+1)^2$$

to give, for every $\epsilon > 0$,

$$\begin{aligned} (4n^2+1)^2 &\ll_\epsilon G(4n^2, (4n^2-1)^2, (4n^2+1)^2)^{1+\epsilon} \\ &= G(4n, 2n-1, 2n+1, 4n^2+1)^{1+\epsilon} \\ &\leq (\sqrt{(2n)(4n^2-1)}(4n^2+1))^{1+\epsilon} \\ &\leq (2n)^{1/2+\epsilon/2}(4n^2+1)^{3/2+3\epsilon/2} \end{aligned}$$

Which, on rearranging, gives a bound on n when ϵ is chosen small.

We end this section with an application of the *abc*-conjecture to an apparently obscure equation. In subsequent sections, the relevance of this equation will become clear.

Proposition 4 *The abc-conjecture implies that there exist only finitely many solutions to the equation $y^3 = x^2 + 1728$ in rational integers.*

Proof: We apply the conjecture to the equation, and find that in any solution (x, y) we must have

$$y^3 \ll (\prod_{p|6xy} p)^{1+\epsilon} \ll (xy)^{1+\epsilon} \ll y^{5(1+\epsilon)/2}.$$

If $\epsilon < 1/5$, this gives a bound on y.

In fact, this equation is known to have only finitely many integral solutions unconditionally - results of Baker and of Stark give estimates for maximum absolute value of integral solutions to the equation $y^3 = x^2 + D$ for $D \in \mathbb{Z} \setminus \{0\}$. In particular a theorem of Stark (see [27, Chapter IX, Section 7]) states that in any integral solution to the above equation we have $\log \max\{|x|, |y|\} \leq C_\epsilon |D|^{1+\epsilon}$ for any $\epsilon > 0$, where C_ϵ is an effectively computable constant. However, the bound on solution size implied by the *abc*-conjecture is considerably better than this bound - the proof of proposition 3 shows that the *abc*-conjecture implies the bound $y^{1-\epsilon} \ll D^{2+\epsilon}$ for all $\epsilon > 0$.

Chapter 2

Background Material

2.1 Elliptic and Modular Functions

2.1.1 Elliptic Functions

Many proofs in this section are omitted; they can be found in [15]. We begin with a few definitions. A *lattice* in $\mathbb{C}$ is a set of the form $\Lambda = [\omega_1, \omega_2] = \{n\omega_1 + m\omega_2 : m, n \in \mathbb{Z}\}$, where ω_1 and ω_2 are linearly independent over $\mathbb{R}$. A *fundamental parallelogram* for Λ is $P = P_\Lambda = \{a\omega_1 + b\omega_2 : a, b \in \mathbb{R}, 0 \leq a, b < 1\}$. A meromorphic function $f : \mathbb{C} \to \mathbb{C}$ is said to be ***elliptic*** with respect to the lattice Λ if for every $\omega \in \Lambda$, and $z \in \mathbb{C}$, we have $f(z + \omega) = f(z)$. In the following, the phrase 'elliptic function' will mean 'elliptic function with respect to some fixed lattice Λ'. Sums, products and quotients of elliptic functions are elliptic, and hence the elliptic function with respect to some lattice form a field. The first example of an elliptic function is the following.

Definition: The *Weierstrass $\wp$-function* for the lattice Λ is defined as follows:

$$\wp(z) = \wp(z, \Lambda) = \frac{1}{z^2} + \sum_{\omega \in \Lambda \setminus \{0\}} \left(\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right).$$

This series can be shown to converge absolutely and uniformly on compact subsets of $\mathbb{C} \setminus \Lambda$, and hence defines a meromorphic function with a pole of order two at every point of Λ. Clearly $\wp$ is even. The uniform convergence allows us to differentiate term by term, and we find that

$$\wp'(z) = \frac{-2}{z^3} - \sum_{\omega \in \Lambda \setminus \{0\}} \frac{2}{(z-\omega)^3}.$$

Clearly we have $\wp'(z+\omega) = \wp'(z)$ for all $\omega \in \Lambda$, since the substitution amounts to rearranging the terms of an absolutely convergent series. Consequently $\wp(z+\omega) = \wp(z) + c$ for some constant $c = c(\omega)$; but since $\wp$ is even, putting $z = -\omega/2$ shows that $c = 0$, and hence that $\wp$ is elliptic.

Theorem 2 *The field of elliptic functions is equal to* $\mathbb{C}(\wp, \wp')$.

Thus, not only is $\wp$ the first example of an elliptic function, but it is in some sense the only example. Using the argument principle, one can show that if f is an elliptic function, f has the same number of zeroes as poles counted with multiplicities in the fundamental parallelogram P. Applying this to $\wp'$ shows that $\wp'$ must have exactly three roots in P. Also, if $u \in \{\frac{\omega_1}{2}, \frac{\omega_2}{2}, \frac{\omega_1+\omega_2}{2}\}$ (the points of order two in $\mathbb{C}/\Lambda$), then $u \equiv -u \pmod{\Lambda}$, and thus $\wp'(u) = 0$, since $\wp'$ is odd and elliptic. Since these three elements are distinct modulo Λ, they must be the only three zeroes of $\wp'$ in P. Now $\wp$ is an elliptic function with two poles in P, and consequently $\wp$ must take any complex value exactly twice in P, since $\wp - c$ for $c \in \mathbb{C}$ must have exactly two zeroes with multiplicity. Therefore for u as above, $\wp$ takes the value $\wp(u)$ only once in P, since these values are taken with multiplicity two. This shows that

$$\frac{(\wp')^2}{(\wp - \wp(\frac{\omega_1}{2}))(\wp - \wp(\frac{\omega_2}{2}))(\wp - \wp(\frac{\omega_1+\omega_2}{2}))}$$

is an elliptic function with no poles on P, which is therefore bounded on $\mathbb{C}$ and hence constant, by Liouville's theorem. A comparison of the $\frac{1}{z^6}$ terms of the numerator and denominator show that this function is in fact identically 4, so $\wp'$ satisfies the differential equation

$$(\wp')^2 = 4\left(\wp - \wp\left(\frac{\omega_1}{2}\right)\right)\left(\wp - \wp\left(\frac{\omega_2}{2}\right)\right)\left(\wp - \wp\left(\frac{\omega_1+\omega_2}{2}\right)\right).$$

Since the three values $\wp(\frac{\omega_1}{2}), \wp(\frac{\omega_2}{2}), \wp(\frac{\omega_1+\omega_2}{2})$ are distinct, this cubic equation has non-zero discriminant. By examining the Laurent series expansion for $\wp'$ and $\wp$ at zero, it can be shown that this cubic is in fact $f(x) = 4x^3 - g_2 x - g_3$, where $g_2 = 60 s_4$, $g_4 = 140 s_6$, and $s_{2k}, k \geq 2$, is defined by

$$s_{2k} = s_{2k}(\Lambda) = \sum_{\omega \in \Lambda \setminus \{0\}} \frac{1}{\omega^{2k}}.$$

Thus $z \mapsto (\wp, \wp')$ defines an analytic map from the Riemann surface $\mathbb{C}/\Lambda$ to an elliptic curve $E = E_\Lambda$ defined by the equation $y^2 = f(x)$. It can be shown that this map is an isomorphism, not only of Riemann surfaces, but of abelian groups, and that furthermore for every elliptic curve over $\mathbb{C}$, there exists a lattice Λ such that $\mathbb{C}/\Lambda \cong E$ (see [27]). We define $\Delta = \Delta_\Lambda = g_2^3 - 27 g_3^2$, so that Δ is 16 times the discriminant of $f(x)$, and is consequently nonzero since $f(x)$ has distinct roots. Note for later reference that the functions s_{2k} satisfy $s_{2k}(c\Lambda) = c^{-2k} s_{2k}(\Lambda)$, and that consequently Δ satisfies $\Delta(c\Lambda) = c^{-12}\Delta(\Lambda)$.

Proposition 1 *If $\mathbb{C}/\Lambda_1 \cong E_1$ and $\mathbb{C}/\Lambda_2 \cong E_2$, and $\phi : E_1 \to E_2$ is an algebraic homomorphism, then there exists $\alpha \in \mathbb{C}$ such that ϕ is induced by multiplication by $\alpha : \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2$. Conversely any such multiplication gives rise to an algebraic homomorphism $E_1 \to E_2$. Furthermore, any analytic map $\mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2$ is in fact induced by multiplication by some $\alpha \in \mathbb{C}$.*

Now suppose that $\Lambda_1 = c\Lambda_2$ for some $c \in \mathbb{C}$. Then multiplication by c induces an isomorphism $E_1 \cong E_2$, and conversely if E_1 and E_2 are isomorphic, their corresponding lattices are multiples of one another, so that isomorphism classes of elliptic curves (over $\mathbb{C}$) are exactly parameterized by lattices in $\mathbb{C}$ modulo nonzero scalar multiplication. This, combined with the above observation regarding the behaviour of the series s_{2k} under scalar multiplication of lattices, provides motivation for the following

Definition: For any lattice Λ in $\mathbb{C}$,

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

The preceding paragraph shows that j is invariant under scalar multiplication of lattices, and consequently that if we define j as a function on elliptic curves by $j(E) = j(\Lambda)$ if $\mathbb{C}/\Lambda \cong E$ that j is invariant on $\mathbb{C}$-isomorphism classes of elliptic curves. In the next section we will show that in fact j parameterizes $\mathbb{C}$-isomorphism classes of elliptic curves, that is that two elliptic curves are isomorphic over $\mathbb{C}$ iff their j values are equal.

2.1.2 Modular Functions

If $\mathcal{L}$ denotes the set of lattices in $\mathbb{C}$, then the functions g_2, g_4, Δ, j defined above are all examples of homogeneous functions $\mathcal{L} \to \mathbb{C}$. Functions such as these are closely associated to functions on the upper half plane $\mathcal{H} = \{z \in \mathbb{C} : \Im(z) > 0\}$ satisfying a certain transformation property. First, since

$$\Im\left(\frac{az+b}{cz+d}\right) = \frac{\Im(z)}{|cz+d|^2},$$

the assignment $z \mapsto \frac{az+b}{cz+d}$ gives an action of $SL_2(\mathbb{Z})$ on $\mathcal{H}$. We denote this by $z \mapsto \gamma(z)$ for $\gamma \in \Gamma$. Note that the matrix $-I$ acts trivially on $\mathcal{H}$. For this reason, we define $\Gamma = SL_2(\mathbb{Z})/\{\pm 1\}$. We will generally denote an arbitrary element of Γ by

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix};$$

suppressing the fact that an element of Γ is in fact a two element equivalence class.

Proposition 2 Γ *is generated by the two matrices*

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

and a fundamental region for the action of Γ *on* $\mathcal{H}$ *is given by* $\mathcal{R} = \{z \in \mathcal{H} : \frac{-1}{2} \leq \Re(z) \leq \frac{1}{2},\ |z| \geq 1\}$.

(By *fundamental region*, we mean that points in the interior of $\mathcal{R}$ are equivalent to no other point of $\mathcal{R}$, and points on the boundary of $\mathcal{R}$ are equivalent to (at most) one other point on the boundary of $\mathcal{R}$.)

Remark - If $\Lambda = [\omega_1, \omega_2]$ is a lattice, we can assume that $\omega_1/\omega_2 \in \mathcal{H}$. This quotient is obviously invariant under scalar multiplication of Λ, and applying an element $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ to ω_1/ω_2 gives $(a\omega_1 + b)/(c\omega_2 + d)$ which also corresponds to Λ. Thus $\Gamma \setminus \mathcal{H}$ is in one to one correspondence with lattices modulo non-zero scalar multiplication, and hence by the above with elliptic curves over $\mathbb{C}$.

Given $F : \mathcal{L} \to \mathbb{C}$ homogeneous of degree $-2k$, we define $f : \mathcal{H} \to \mathbb{C}$ as follows: $f(z) = F(\Lambda_z)$, where Λ_z is the lattice $[z, 1]$. For any $\gamma \in \Gamma$, and any lattice $\Lambda = [\omega_1, \omega_2]$, $\gamma(\Lambda) = [a\omega_1 + b\omega_2, c\omega_1 + d\omega_2] = \Lambda$, and hence $(cz + d)\Lambda_{\gamma(z)} = \Lambda_z$. In terms of f, this means that $f(\gamma z) = F\left(\frac{1}{cz+d}\Lambda_z\right) = (cz + d)^{2k} f(z)$. This gives us an association between homogeneous lattice functions of degree $-2k$ and functions f on $\mathcal{H}$ satisfying

$$(2.1) \qquad f(\gamma z) = (cz + d)^{2k} f(z) \text{ for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma.$$

Definition: A meromorphic function satisfying (2.1) is called a *modular function of weight* $2k$.

Note that since the action of $T \in \Gamma$ on $\mathcal{H}$ is $z \mapsto z+1$, a modular function (of any weight) satisfies $f(z + 1) = f(z)$, and hence if $\mathbb{D}^* = \{z \in \mathbb{C}^* : |z| < 1\}$ and exp denotes the exponential function $\mathcal{H} \to \mathbb{C}$ defined by $\exp(z) = e^z$ then f defines a meromorphic function f^* on $\mathbb{D}^*$, given by $f^*(\exp(2\pi iz)) = f(z)$. This function has a Laurent series expansion in $q = \exp(2\pi iz)$ about $q = 0$, and if this series has no nonzero negative power coefficients (that is, if f^* is analytic on $\mathbb{D} = \mathbb{D}^* \bigcup\{0\})$ and f is holomorphic on $\mathcal{H}$ then f is said to be a *modular form.* Thus we are defining modular forms to be holomorphic functions on the Riemann surface $\Gamma \setminus \mathcal{H}$, where the differentiable structure at ∞ is given via the map $z \mapsto \exp(z)$. This statement, together with the remark after proposition 2 give a reason behind the importance of modular forms for the study of elliptic curves. If a modular form satisfies the additional condition $f^*(0) = 0$ then it is said to be a *cusp form*. The q-expansion of f^* is said to be the expansion of f *at infinity*, and if f is a modular form we define $f(\infty) = f^*(0)$. Now beginning with the homogeneous lattice functions s_{2k} ($k \geq 2$) introduced above, we define the corresponding modular functions,

$$G_{2k} = \sum_{m,n \in Z}{}' \frac{1}{(mz+n)^{2k}},$$

where the prime on the sum means that the $m = n = 0$ term is left out of the sum; and as before we put $g_2 = 60G_4$, $g_3 = 140G_6$. Note that we are using g_2 and g_3 to denote both lattice functions and the corresponding modular functions. We will also denote by Δ and j respectively both the lattice functions defined above and the corresponding modular functions. Since the series defining G_{2k} ($k \geq 2$) converge absolutely and uniformly on compact subsets of $\mathcal{H}$, the functions G_{2k} are holomorphic on $\mathcal{H}$, and putting '$z = \infty$' formally into the above series gives $G_{2k}(\infty) = 2\zeta(2k)$, where $\zeta(z)$ is the Riemann zeta function. The q-expansion of G_{2k} at infinity does indeed have constant term $2\zeta(2k)$ (see below for the q-expansions of g_2 and g_3), so this formal substitution is justified, and using the facts that $\zeta(4) = \pi^4/90$ and $\zeta(6) = \pi^6/945$, an easy calculation shows that $\Delta(\infty) = 0$, so that Δ is a cusp form of weight 12. Since $\Delta(z) \neq 0$ for all $z \in \mathcal{H}$, we see that j is holomorphic on $\mathcal{H}$. (This explains the presence of the factors of 60 and 140 in the definitions of g_2 and g_3; the 1728 in the definition of j is so that j has a q-expansion at ∞ with all integer coefficients; see section 2.1.3.)

If f is meromorphic in a neighbourhood of the point a, we define the *order of* f at a, $v_a(f)$, to be the unique integer m such that $(z-a)^{-m}f$ is defined and non-zero at a (the existence and uniqueness of this integer follow from the fact that f has a Laurent series expansion about a with only finitely many non-zero negative power coefficients). With this notation, we have the following formula, proved by integrating the function f'/f around the boundary of a fundamental domain for $\Gamma \setminus \mathcal{H}$.

Proposition 3 *If f is a nonzero modular function of weight 2k, then*

$$v_\infty(f) + \frac{v_\rho(f)}{3} + \frac{v_i(f)}{2} + \sum_{p\neq i,\rho} v_p(f) = \frac{k}{6} \tag{2.2}$$

where the sum is over all p in a fundamental domain for $\Gamma \setminus \mathcal{H}$, and $\rho = e^{\frac{\pi i}{3}}$ is a primitive third root of unity.

This formula allows us to prove the following three propositions:

Proposition 4 *The zero of Δ at ∞ is simple.*

Proof: This follows immediately from the above formula, since for Δ, a cusp form of weight 12, the right hand side is 1 and the left hand side is equal to the multiplicity of the zero at ∞, since Δ is holomorphic and non-vanishing on $\mathcal{H}$.

Proposition 5 *The map $j : \Gamma \setminus \mathcal{H} \to \mathbb{C}$ is a bijection.*

Proof: For any $c \in \mathbb{C}$, the function $j - c$ is modular of weight zero, with a simple pole at infinity. Consequently (2.2) gives

$$\frac{v_\rho(f)}{3} + \frac{v_i(f)}{2} + \sum_{p\neq i,\rho} v_p(f) = 1.$$

Since j is analytic on $\mathcal{H}$ all the terms on the left hand side are positive, and since 2 and 3 are relatively prime, the equation can only be satisfied if exactly one term is nonzero. Thus the map is indeed a bijection.

Remark - This shows that if Λ_1 and Λ_2 are $\mathbb{C}$-lattices then $j(\Lambda_1) = j(\Lambda_2)$ iff $\Lambda_1 = c\Lambda_2$ for some $c \in \mathbb{C}$, and consequently provides a proof of the earlier assertion that the j-function parameterizes $\mathbb{C}$-isomorphism classes of elliptic curves.

Note that j takes the values $j(\rho)$ and $j(i)$ with multiplicities 3 and two respectively. The reason for this will be seen in the proof of the next proposition.

Proposition 6 *Let M_k (resp. S_k) denote the $\mathbb{C}$-vector space of modular forms (resp. cusp forms) of weight 2k. Then:*

(a) $\dim(M_0) = 1$

(b) $\dim(M_1) = 0$

(c) $\dim(M_k) = 1$ *if* $k = 2, 3, 4$ *or* 5

(d) $\dim(M_k) = \dim(M_{k-6}) + 1$ *if* $k \geq 6$

(e) $\dim(S_k) = \dim(M_k) - 1$

Proof: If $f \in M_k$, f has no poles, so all terms on the left hand side of (2.2) are nonnegative. If $f \in M_0$, $f \neq 0$, then f has no zeroes by (2.2). Now the constant functions are in M_0, and so for any $z \in \mathcal{H}$, $f - f(z)$ is in M_0, and has a zero. Hence $f - f(z) \equiv 0$, so f is constant, proving (a). For (b), note that if f is a nonzero element of M_1, the equation (2.2) cannot hold, since the left hand side is 0 or $\geq 1/3$, and the right hand side is 1/6. Now (c) is true since $M_2 = \mathbb{C}g_2$, $M_3 = \mathbb{C}g_3$, $M_4 = \mathbb{C}g_2^2$, $M_5 = g_2g_3$. We'll prove the first of these statements; the proofs of the others are identical. Let $f \in M_2$, then the left hand side of (2.2) is 1/3, so f must have a simple zero at ρ and no others. Take $x \in \mathcal{H}$, $x \neq \rho$, and consider the function $g_2(x)f - f(x)g_2$. This function has zeroes at both ρ and x and hence must be zero everywhere. Thus $f \in \mathbb{C}g_2$, proving that $M_2 = \mathbb{C}g_2$.

If $k \geq 6$, let $f \in M_k$. then there exists a $c \in \mathbb{C}$ such that $f - cG_{2k} \in S_k$; therefore

$$\frac{f - cG_{2k}}{\Delta} \in M_{k-6}.$$

Hence the map $h \mapsto h\Delta$ defines an isomorphism $S_k \cong M_{k-6}$, so

$$M_k = S_k \bigoplus \mathbb{C}G_{2k},$$

proving (d) and (e).

Note that g_2 has a simple zero at ρ and no other zeroes; the proof that $M_3 = \mathbb{C}g_3$ shows that g_3 has a simple zero at i and no other zeroes, and differentiating j using the quotient rule shows that zeroes of j' correspond to zeroes of g_2, zeroes of g_3, or points where $3g_3g_2' - 2g_2g_3' = 0$. So the zeroes of g_2 and g_3 at ρ and i respectively explain the double and triple zeroes of j, and since the proof of the previous proposition showed that j could have non-simple zeroes exactly at ρ and i, we conclude that there can exist no points of $\mathcal{H}$ where $3g_3g_2' - 2g_2g_3' = 0$.

Corollary *Up to multiplication by elements of $\mathbb{C}$, j is the only weight zero modular function holomorphic on $\mathcal{H}$ with a simple pole at ∞.*

Proof: j has a simple pole at ∞ since Δ has a simple zero there, and for any f satisfying the hypotheses, $f\Delta \in M_6 \setminus S_6 = \mathbb{C}G_{12}$ and hence $f\Delta$ is a $\mathbb{C}$-multiple of $j\Delta$, so f is a $\mathbb{C}$-multiple of j.

2.1.3 Some q-Expansions

In this section we derive a few important q-expansions, largely in order to prove that j has a q-expansion with integral coefficients.

Proposition 7 *The functions G_1, g_2, and g_3 have the following q-expansions:*

$$G_1 = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty}{}' \frac{1}{(m+nz)^2} = \frac{-\pi}{3} - 8\pi^2 \sum_{n=1}^{\infty} \sigma_1(n) q^n$$

$$g_2 = 60 \sum_{m,n \in \mathbb{Z}}{}' \frac{1}{(m+nz)^4} = \frac{4\pi^3}{3}\left(1 + 240 \sum_{k=1}^{\infty} \sigma_3(k) q^k\right)$$

$$g_3 = 140 \sum_{m,n \in \mathbb{Z}}{}' \frac{1}{(m+nz)^6} = \frac{8\pi^6}{27}\left(1 - 504 \sum_{k=1}^{\infty} \sigma_5(k) q^k\right),$$

where $\sigma_n(k) = \sum_{d|k} d^n$. Note that G_1 is not an absolutely convergent series, and consequently the order of summation is important here.

Proof: We prove the formula for g_2, the other derivations are similar. The definition of g_2 gives immediately

$$g_2(z) = 120\zeta(4) + 60 \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} \frac{1}{(m+nz)^4} + \frac{1}{(m-nz)^4},$$

and on replacing $-n$ by n we have

$$(2.3) \qquad g_2(z) = 120\zeta(4) + 120 \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} \frac{1}{(m+nz)^4}.$$

Now consider the factorization of the sine function given by the Weierstrass factorization theorem:

$$\sin(\pi z) = \pi z \prod_{m=1}^{\infty} \left(1 - \left(\frac{z}{m}\right)^2\right).$$

Differentiating this factorization logarithmically gives

$$(2.4) \qquad \pi \cot(\pi z) = \frac{1}{z} + \sum_{m=1}^{\infty}{}' \frac{1}{z+m} + \frac{1}{z-m} = \frac{1}{z} + \sum_{m=1}^{\infty} \frac{2z}{z^2-m^2}.$$

Now for any $z \in \mathcal{H}$, $|q| = |e^{2\pi iz}| < 1$, so

$$\pi \cot(\pi z) = \pi i \frac{q+1}{q-1} == -\pi i(q+1) \sum_{k=0}^{\infty} q^k = -\pi i \left(1 + 2\sum_{k=1}^{\infty} q^k\right).$$

Comparing the right hand sides of the last two equations gives

$$(2.5) \qquad \frac{1}{z} + \sum_{m=1}^{\infty} \frac{2z}{z^2-m^2} = -\pi i \left(1 + 2\sum_{k=1}^{\infty} q^k\right),$$

and differentiating three times yields the equation

$$(2.6) \qquad \left(\sum_{m=-\infty}^{\infty} \frac{1}{(z+m)^4}\right) = \frac{(2\pi i)^4}{6} \left(\sum_{k=1}^{\infty} k^3 q^k\right).$$

Now inserting (2.6) into (2.3) with z replaced by nz yields

$$g_2(z) = 120 \left(\zeta(4) + \frac{8\pi^4}{3} \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} k^3 q^{nk}\right).$$

Now for a given power l of q in this equation, the coefficient is $\sigma_3(l)$, since q^l appears in the n sum exactly once for each $d|l$, with coefficient $(\frac{n}{d})^3$. This proves the proposition.

Note that (2.4) and (2.5) above, when differentiated once, yield

$$\sum_{n=-\infty}^{\infty} \frac{1}{(z+n)^2} = (2\pi i)^2 \sum_{n=1}^{\infty} nq^n,$$

so for z, $\tau \in \mathcal{H}$, we have (putting $r = e^{2\pi i \tau}$)

$$\begin{aligned}
\wp(z;[\tau,1]) &= \frac{1}{z^2} + \sum_{m,n\in \mathrm{Z}}{}' \left(\frac{1}{(z-m\tau-n)^2} - \frac{1}{(m\tau+n)^2} \right) \\
&= \frac{1}{z^2} + \sum_{n=-\infty}^{\infty}{}' \left(\frac{1}{(z-n)^2} - \frac{1}{n^2} \right) + \\
&\quad \sum_{m=1}^{\infty} \left[\sum_{n\in \mathrm{Z}} \left(\frac{1}{(z-m\tau-n)^2} + \frac{1}{(-z+m\tau+n)^2} \right) - 2\sum_{n\in \mathrm{Z}} \frac{1}{(m\tau+n)^2} \right] \\
&= (2\pi i)^2 \sum_{n=1}^{\infty} nq^n - 2\zeta(2) + (2\pi i)^2 \sum_{m=1}^{\infty} \left[\sum_{n=1}^{\infty} (n(qr^m)^n + n(q^{-1}r^m)^n) - 2\sum_{n=1}^{\infty} n(r^m)^n \right] \\
&= -4\pi^2 \left[\frac{1}{12} + \sum_{n=1}^{\infty} nq^n + \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} nr^{mn}(q^n+q^{-n}) - 2\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} nr^{mn} \right]
\end{aligned}$$

which gives us a (q,r)-expansion of $\wp$ with integral coefficients (up to a multiple of $4\pi^2$). This will be important later.

Proposition 8 $\Delta = (2\pi)^{12} q \prod_{n=1}^{\infty} (1-q^n)^{24}$.

Proof: (This proof is due to Hurwitz, and appears in Chapter two of [26]) Put $F(z) = q\prod_{n=1}^{\infty}(1-q^n)^{24}$. Then F represents a holomorphic function on $\mathcal{H}$, since $z \in \mathcal{H}$ means that $|q| < 1$, so the product for F converges uniformly and absolutely on compact subsets of $\mathcal{H}$. Furthermore it is clear that $F(\infty) = 0$, so since we know that $\Delta \in S_6$ which is a one-dimensional $\mathbb{C}$-vector space, to prove F and Δ are proportional, it suffices to show that F satisfies the transformation laws for a level 12 modular form. $F(z+1) = F(z)$ from the q-product given, and we now prove that $F(\frac{-1}{z}) = z^{12}F(z)$. We define the following series:

$$G(z) = \sum_m \sum_n{}' \frac{1}{(m+nz)^2} \qquad G_1(z) = \sum_n \sum_m{}' \frac{1}{(m+nz)^2}$$

$$H(z) = \sum_m \sum_n{}' \frac{1}{(m-1+nz)(m+nz)}$$

$$H_1(z) = \sum_n \sum_m{}' \frac{1}{(m-1+nz)(m+nz)},$$

where the $'$ in the G and G_1 sums means to leave off the term $(m,n)=(0,0)$ and the $'$ in the H and H_1 sums means to leave off $(m,n)=(0,0)$ and $(m,n)=(1,0)$, that is $'$ means to leave off any terms which would cause the series to be undefined. The H and H_1 sums are essentially tools to allow us to evaluate the difference $G_1 - G$. In the definition of H_1, for any $n \neq 0$, the sum telescopes to zero:

$$\sum_{m=-\infty}^{\infty} \frac{1}{(m-1+nz)(m+nz)} = \sum_{m=-\infty}^{\infty} \left(\frac{1}{(m-1+nz)} - \frac{1}{m+nz} \right) = 0.$$

Thus,

$$H_1(z) = \sum_{m \neq 0,2} \left(\frac{1}{m-1} - \frac{1}{m} \right) = 2,$$

since this sum is also telescopic and the only terms which survive are from $m = 0, 2$, which both give a contribution of 1. To evaluate H, we return to the expansion of the cotangent function used in the proof of the last proposition, and replace z by m/z, to yield

$$\begin{aligned}
\frac{\pi}{z} \cot\left(\frac{\pi m}{z}\right) &= \frac{1}{m} + \sum_{n=1}^{\infty} \left(\frac{1}{m+nz} + \frac{1}{m-nz} \right) \\
&= \frac{1}{m} \sum_{n=1}^{\infty} \left(\frac{1}{m+nz} - \frac{1}{nz} \right) + \left(\frac{1}{m-nz} - \frac{1}{(-n)z} \right) \\
&= \frac{1}{m} + \sum_{-\infty}^{\infty}{}' \left(\frac{1}{m+nz} - \frac{1}{nz} \right),
\end{aligned}$$

since

$$\sum_{n=1}^{\infty} \left(\frac{1}{m+nz} - \frac{1}{nz} \right) = \sum_{n=1}^{\infty} \frac{-m}{nz(nz+m)}$$

converges absolutely. Hence for any $m \neq 0, 1$ the sum on n in H_1 is

$$\sum_{n=-\infty}^{\infty} \left(\frac{1}{(m-1+nz)} - \frac{1}{(m+nz)} \right) = \frac{\pi}{z} \left(\cot\left(\frac{\pi(m-1)}{z} \right) - \cot\left(\frac{\pi m}{z} \right) \right),$$

While the $m = 0$ sum gives

$$-\sum_{n=-\infty}^{\infty}{}' \left(\frac{1}{1+nz} - \frac{1}{nz}\right) = 1 - \frac{\pi}{z}\cot\left(\frac{\pi}{z}\right),$$

and the $m = 1$ sum gives the same contribution. Summing all terms except for $m = 0, 1$ yields

$$\begin{aligned} &\lim_{M\to\infty} \frac{\pi}{z} \sum_{m=-M}^{M} \left(\cot\left(\frac{\pi(m-1)}{z}\right) - \cot\left(\frac{\pi m}{z}\right)\right) \\ = \; &\lim_{M\to\infty} \frac{\pi}{z} \left(\cot\left(\frac{\pi(-M-1)}{z}\right) - \cot\left(\frac{\pi(M)}{z}\right) - 2\cot\left(\frac{\pi}{z}\right)\right) \\ = \; &\frac{-2\pi i}{z} - \frac{2\pi}{z}\cot\left(\frac{\pi}{z}\right), \end{aligned}$$

since $\lim_{M\to\infty} \cot\left(\frac{M\pi}{z}\right) = i$ for any $z \in \mathcal{H}$. Thus adding all the terms shows that $H(z) = 2 - 2\pi i/z$, so that $H - H_1 = -2\pi i/z$.

Now the series

$$H - G = \sum_m \sum_n{}' \left(\frac{1}{(m-1+nz)(m+nz)} - \frac{1}{(m+nz)^2}\right)$$

converges absolutely, and therefore is independent of the ordering of the summands. This tells us that $H - G = H_1 - G_1$, and thus that $G - G_1 = H - H_1 = -2\pi i/z$. Furthermore, $G_1(-1/z) = z^2 G(z)$, so $G_1(-1/z) = z^2 G_1(z) - 2\pi i z$. Now taking the logarithmic derivative of F we have

$$\frac{F'}{F} = 2\pi i \left(1 - 24\sum_{k=1}^{\infty} \sigma_1(k) q^k\right) dz,$$

and combined with the q-expansion of G_1 given in the previous proposition we have

$$\frac{F'}{F} = \frac{6i}{\pi} G_1(z) dz.$$

Now taking the logarithmic derivative of $F\left(\frac{-1}{z}\right)$ shows that

$$\frac{d(F(-1/z))}{F(-1/z)} = \frac{6i}{\pi} G_1(z) dz + \frac{12dz}{z} = \frac{d(z^{12}F(z))}{z^{12}F(z)},$$

so that $F(-1/z)$ and $z^{12}F(z)$ have the same logarithmic derivative. Hence there exists some $c \in \mathbb{C}$ such that $F(-1/z) = cz^{12}F(z)$. For $z = i$, $z = -1/z$ and $z^{12} = 1$ and $F(z) \neq 0$ (since F is non-vanishing on $\mathcal{H}$) so $c = 1$. This shows that F is a constant multiple of Δ. Since the q term of the expansion of F is clearly one, the value of this constant will be the q-term of the q-expansion of $\Delta = g_2^3 - 27g_3^2$, which by the above proposition is

$$\left(\frac{4\pi^4}{3}\right)^3 (720) - 27\left(\frac{8\pi^6}{27}\right)^2 (-1008) = (2\pi)^{12}.$$

We now justify the factor of 1728 in the definition of j.

Theorem 3

$$j(z) = \frac{1}{q} + 744 + \sum_{n=1}^{\infty} c_n q^n,$$

with each $c_n \in \mathbb{Z}$.

Proof: Using the above formulas we have

$$g_2^3 = \frac{64\pi^{12}}{27}(1 + 720q + \cdots),$$

$$\frac{\Delta(z)}{1728} = \frac{64\pi^{12}}{27} q (1 - 24q + \cdots)$$

where the $\cdots$ represent power series in q with integral coefficients. Now the series $1 - 24q + \cdots$ has leading coefficient a unit of $\mathbb{Z}$ and is thus invertible in the power series ring $\mathbb{Z}[[q]]$, with inverse $1 + 24q + \cdots$. Since the function Δ is holomorphic and nonvanishing on $\mathcal{H}$, this inverse is holomorphic, and putting the above two equations into the definition of j we find that

$$j(z) = \frac{1728g_2^3}{\Delta} = \frac{1}{q}(1 + 720q + \cdots)(1 + 24q + \cdots) = \frac{1}{q} + 744 + \sum_{n=1}^{\infty} c_n q^n$$

with the $c_n \in \mathbb{Z}$, as required.

2.2 Modular Function Fields

In many situations it is useful to consider functions which satisfy the modular functional equation

$$f(\gamma z) = (cz + d)^{2k} f(z)$$

for all γ in some specified subgroup of Γ. In particular, we define a family of normal subgroups of Γ as follows:

$$\Gamma(N) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ (mod\, N)\}$$

($\Gamma(N)$ is normal since it is the kernel of the reduction map $\Gamma \rightarrow SL_2(\mathbb{Z}/N\mathbb{Z})$). A meromorphic function f satisfying the above equation for all $\gamma \in \Gamma(N)$ is called a *modular function of weight* $2k$ *and level* N if it is also 'meromorphic at the cusps', that is if for all $\gamma \in \Gamma$, $f(\gamma z)$ has a power series expansion at ∞ in powers of $q^{1/N}$ with only finitely many non-zero negative power coefficients. Since the product and quotient of two nonzero weight zero modular functions of any level is another weight zero modular function of the same level, the set of all modular functions of weight zero and level N forms a field which we will denote by $\mathcal{F}_N$.

Examples of modular functions of weight zero and level N are given by the Fricke functions, which we define in the following manner. Let Λ be a lattice, and $\wp(z) = \wp(z; \Lambda)$ the corresponding Weierstrass function. Assume that $g_2(\Lambda) \neq 0 \neq g_3(\Lambda)$. Define the *Weber function* f_0 by

$$f_0(z, \Lambda) = \frac{g_2(\Lambda) g_3(\Lambda)}{\Delta(\Lambda)} \wp(z; \Lambda).$$

Since $\wp(cz; c\Lambda) = c^{-2}\wp(z; \Lambda)$, using the homogeneity properties of g_2, g_3 and Δ, we see that the function f_0 is invariant under isomorphisms $\mathbb{C}/\Lambda \rightarrow \mathbb{C}/c\Lambda$. Now we define the *Fricke functions* $f_{a,b}$ by

$$f_{a,b}(z) = f_0(\frac{az + b}{N}; \Lambda_z),$$

where $\Lambda_z = [z, 1]$. Note that if the normalizing factor $g_2 g_3/\Delta$ is left off, the values of the functions $f_{u,v}(z)$ are the x-coordinates of the N-division points of the elliptic curve corresponding to the lattice Λ_z. The values $f_{u,v}(z)$, therefore, while not necessarily corresponding to the N-division points of an elliptic curve, can be thought of as corresponding to 'normalized' N-division points of any elliptic curve isomorphic to $\mathbb{C}/\Lambda_z$, *invariant under automorphisms.* These functions only depend on a and b up to congruence modulo N, since $\wp$ is elliptic. This means that each $f_{u,v}$ is invariant under $\Gamma(N)$: If

$$\gamma = \begin{pmatrix} 1+Na & Nb \\ Nc & 1+Nd \end{pmatrix} \in \Gamma(N),$$

then for each Fricke function we have

$$\begin{aligned} f_{u,v}(\gamma z) &= f_0\left(\frac{u \begin{pmatrix} 1+Na & Nb \\ Nc & 1+Nd \end{pmatrix} z + v}{N}; \ \Lambda_{\gamma z} \right) \\ &= f_0\left(\frac{u[(1+Na)z+Nb]+v[Ncz+(1+Nd)]}{(N)(Ncz+(1+Nd))}; \ \frac{1}{Ncz+(1+Nd)}\Lambda_z \right) \\ &= f_0\left(\frac{uz+b}{N}; \ \Lambda_z \right), \end{aligned}$$

with the third equality coming because of the degree zero homogeneity (that is, isomorphism invariance) of f_0 and the *mod* N periodicity of the $f_{u,v}$. Furthermore, the q-expansions of the previous section imply that the functions $f_{u,v}$ have $q^{1/N}$-expansions with only finitely many non-zero negative power coefficients, and that furthermore all the coefficients of the q-expansion lie in $\mathbb{Q}(\zeta_n)$ (ζ_n being a primitive N^{th} root of unity). Hence each $f_{u,v} \in \mathcal{F}_N$. In fact,

Theorem 4 *(a)* $\mathcal{F}_1 = \mathbb{C}(j)$.

(b) Let $K = \mathbb{C}(j, \{f_{a,b} : (a,b) \in \frac{\mathbb{Z}}{N\mathbb{Z}}\})$*. Then* $K = \mathcal{F}_N$*, and* $Gal(\mathcal{F}_N/\mathbb{C}(j)) = \Gamma/\Gamma(N) = SL_2(\mathbb{Z}/N\mathbb{Z})$.

Proof: For (a), take $f \in \mathcal{F}_1$ and let $a_1, ..., a_n$ be the poles of f, counted with multiplicities. Then $f(j - j(a_1))...(j - j(a_n))$ is analytic on H and meromorphic at ∞, and is hence in $\mathbb{C}[j]$, so $f \in \mathbb{C}(j)$.

For (b), note that we have a homomorphism $\phi : \Gamma \to Gal(\mathcal{F}_N/\mathcal{F}_1)$, since Γ acts as a group of automorphisms of $\mathcal{F}_N$ by $f \mapsto f \circ \gamma$. The fixed field of this action is exactly $\mathcal{F}_1$ so $\mathcal{F}_N$ is a finite Galois extension of $\mathcal{F}_1$ (finite since $Ker(\phi) \supseteq \Gamma(N)$), and the fundamental theorem of Galois theory states that $\phi(\Gamma) = Gal(\mathcal{F}_N/\mathcal{F}_1)$. Now we consider K. Γ also acts as a group of automorphisms of K, since

$$\begin{aligned} f_{u,v}(\gamma z) &= f_0\left(\frac{u(az+b)+v(cz+d)}{N};\ \Lambda_z\right) \\ &= f_{ua+vc,ub+vd}(z) = f_{(u,v)\gamma}(z), \end{aligned}$$

and thus $K/\mathcal{F}_1$ is finite Galois and we have a homomorphism $\psi : \Gamma \to Gal(K/\mathcal{F}_1)$; again the kernel contains $\Gamma(N)$. We now show that the kernel of ψ is exactly equal to $\Gamma(N)$. Since we know that $K \subseteq \mathcal{F}_N$, This will prove that $K = \mathcal{F}_N$, since it will show that

$$Gal(K/\mathcal{F}_1) \supseteq Im(\psi) = \frac{\Gamma}{Ker(\psi)} = \frac{\Gamma}{\Gamma(N)} \supseteq \frac{\Gamma}{Ker(\phi)} = Gal(\mathcal{F}_N/\mathcal{F}_1).$$

Let $\alpha = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \in \Gamma$, and assume that α acts trivially on K. Then $f_{1,0}(\alpha z) = f_{1,0}(z)$ for all z, so we have

$$\wp\left(\frac{z}{N};\ \Lambda_z\right) = \wp\left(\frac{pz+q}{N};\ \Lambda_z\right);$$

but we know that $\wp(z_1) = \wp(z_2)$ iff $z_1 \equiv \pm z_2\ (mod\ \Lambda)$. This means that $p \equiv \pm 1$ and $q \equiv 0$ modulo Λ. Applying the same argument to $f_{1,0}$ shows that $r \equiv 0$ and $s \equiv \pm 1$ modulo Λ. If $N = 2$, we are finished, and if $N > 2$, the fact that $det(\alpha) = 1$ shows that the p and s must have the same parity modulo N, so that $\alpha \equiv I\ (mod\ N)$ (remember that $\Gamma = SL_2(\mathbb{Z})/\{\pm 1\}$). This completes the proof.

Let $F_N = \mathbb{Q}(j, \{f_{u,v}\})$. We will call this the *field of modular functions over* $\mathbb{Q}$.

Theorem 5 *F_N is a Galois extension of $\mathbb{Q}(j)$ with Galois group $G = GL_2(\mathbb{Z}/N\mathbb{Z})$.*

Proof: The previous theorem implies that the group G in question contains $SL_2(\mathbb{Z}/N\mathbb{Z})$. This is a normal subgroup of $GL_2(\mathbb{Z}/N\mathbb{Z})$, and the elements of the factor set are represented by the values of the determinant possible for elements of $GL_2(\mathbb{Z}/N\mathbb{Z})$, of which there are $\phi(N)$ (Euler ϕ-function). Thus, if we show that G contains an element of each possible determinant, we will know that G contains $GL_2(\mathbb{Z}/N\mathbb{Z})$. The functions $f_{u,v}$ have $q^{1/N}$-expansions with coefficients in $\mathbb{Q}(\zeta_N)$, and are hence elements of $\mathbb{Q}((q^{1/N}))$, so the automorphisms $\zeta_N \mapsto \zeta_N^d$ of $\mathbb{Q}(\zeta_N)$ ($d \in (\mathbb{Z}/N\mathbb{Z})^*$) act on the elements $f_{u,v}$ (action is given by action on Fourier coefficients), and these automorphisms fix $\mathbb{Q}(j)$, since j has Fourier coefficients in $\mathbb{Z}$. A simple calculation shows that these automorphisms in fact send $f_{u,v}$ to $f_{u,dv}$, so these automorphisms can be represented by the matrices $\begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix}$. Thus the Galois group contains $GL_2(\mathbb{Z}/N\mathbb{Z})$. To show the other containment, we consider the polynomial

$$\prod_{u,v} (X - f_{u,v}),$$

where the product is taken over all $(u, v) \in (\mathbb{Z}/N\mathbb{Z})^2$. This polynomial has as X-power coefficients symmetric functions in the $f_{u,v}$, and since the elements of Γ permute the $f_{u,v}$, these coefficients are invariant under Γ. Since they are also holomorphic on $\mathcal{H}$ and have Fourier expansions with coefficients in $\mathbb{Z}(\zeta_N)$, they are polynomials in j with coefficients in $\mathbb{Q}(\zeta_N)$. This shows that the functions $f_{u,v}$ are algebraic over $\mathbb{Q}(j)$, and since Γ permutes the $f_{u,v}$ transitively, the polynomial must be irreducible over $\mathbb{C}(j)$. This shows that

$$\begin{aligned} [F_N : \mathbb{Q}(j)] &\leq [\mathcal{F}_N : \mathbb{C}(j)][\mathbb{Q}(\zeta_N) : \mathbb{Q}] \\ &\leq |SL_2(\mathbb{Z}/N\mathbb{Z})|\,\phi(N) \\ &= |GL_2(\mathbb{Z}/N\mathbb{Z})|, \end{aligned}$$

and hence that $F_N/\mathbb{Q}(j)$ is a Galois extension with group $GL_2(\mathbb{Z}/N\mathbb{Z})$.

2.3 Automorphisms

When we defined the Weber and Fricke functions in the previous section we were assuming that the lattice Λ was such that $g_2(\Lambda)$ and $g_3(\Lambda)$ were both non-zero, that is that Λ was not a multiple of $[\imath, 1]$ or $[\rho, 1]$. These two lattices correspond to special isomorphism classes of elliptic curves over $\mathbb{C}$, those with non-trivial automorphisms, and for these lattices the Weber functions are defined differently. Specifically, if $g_2(\Lambda) = 0$, put

$$f_0(z; \Lambda) = \frac{g_3(\Lambda)}{\Delta(\Lambda)}(\wp(z))^3,$$

and if $g_3(\Lambda) = 0$ put

$$f_0(z; \Lambda) = \frac{g_2^2(\Lambda)}{\Delta(\Lambda)}(\wp(z))^2.$$

Since g_2 is homogeneous of degree -4, g_3 is homogeneous of degree -6 and Δ is homogeneous of degree -12, these functions are both homogeneous of degree zero - invariant under isomorphisms $\frac{\mathbb{C}}{\Lambda} \to \frac{\mathbb{C}}{c\Lambda}$. Furthermore, the Fourier expansions of the last section show that the Weber functions can be defined by (q, r)-series with coefficients in $\mathbb{Q}$. The following theorem shows the importance of the Weber functions.

Theorem 6 *Let Λ be a lattice in $\mathbb{C}$, and let f_0 be the Weber function for Λ, as defined above. Then $f_0(z_1) = f_0(z_2)$ iff there exists some automorphism $\mathbb{C}/\Lambda \to \mathbb{C}/\Lambda$ which takes z_1 to z_2.*

The proof of this theorem is an application of the following lemma:

Lemma 1 *Let $\Lambda = [\alpha, 1]$ be a lattice of $\mathbb{C}$ with $\alpha \in \mathbb{C}$, and suppose there exists some $c \in \mathcal{H}$ such that $c \neq \pm 1$ and $c\Lambda = \Lambda$. Then either $c = \pm i$ and Λ is a multiple of $[i, 1]$ or $c = \pm\rho$ or $c = \pm\rho^2$ and Λ is a multiple of $[\rho, 1]$.*

Proof: First, we show that the existence of a $c \in \mathbb{C}$ such that $c\Lambda \subseteq \Lambda$ shows that α is imaginary quadratic. For such a c, we have $c\alpha = m_1\alpha + n_1$ and $c = m_2\alpha + n_2$ with m_1, m_2, n_1, n_2 all integers. Substituting the former equation into the latter shows that α satisfies a quadratic equation with coefficients in $\mathbb{Z}$, and thus that α is imaginary quadratic, since $\alpha \in \mathcal{H}$. Denote by K the imaginary quadratic field containing α. Then $c \in K$, since $c = m_2\alpha + n_2$. We now assume that $c\Lambda = \Lambda$. $\{\alpha, 1\}$ is a $\mathbb{Q}$-basis of K, and so we have that $1 = [\Lambda : c\Lambda] = N_Q^K(c)$. Thus c is a unit in an imaginary quadratic field, not equal to ± 1, and hence we must have $c = \pm i$ or $c = \pm\rho$ or $c = \pm\rho^2$. There are now three cases to consider, according as which of these numbers c is equal to. We will do only the case $c = \pm\rho^2$, the others are treated identically. $c = \rho^2 = -\rho - 1$ implies that $K = \mathbb{Q}(\sqrt{-3})$, and we must have $\alpha = a\rho + b$ for $a, b \in \mathbb{Q}$. We have that $[a + b(-\rho - 1), (-\rho - 1)] = [a\rho + b, 1]$, and calculation shows that this can happen only if $(a, b) = (1, 0)$, $(1, 1)$, $(1, -1)$, or $(1/3, -1/3)$. The first three cases all lead to the conclusion that $[a\rho + b, 1] = [\rho, 1]$, while the fourth leads to the conclusion that $[a\rho + b, 1] = (\sqrt{-3})[\rho, 1]$. (In the case $c = i$, we find that $\Lambda = [i, 1]$ or $\Lambda = (1 + i)[i, 1]$.)

Note that the in the proof of the above lemma it was shown that if $c\Lambda \subseteq \Lambda$ for some lattice Λ, then c is imaginary quadratic. Using the equivalence of algebraic maps of elliptic curves with functions on lattices (see section 2.2.1), this shows that any endomorphism of an elliptic curve over $\mathbb{C}$ can be identified with an imaginary quadratic complex number. The set of all endomorphisms of the elliptic curve $E/\mathbb{C}$ form a ring called $End(E)$; this ring is either $\mathbb{Z}$ or an order of an imaginary quadratic field (in the latter case, the curve E is said to have *complex multiplication*).

We now prove the theorem. If Λ is neither a multiple of $[\rho, 1]$ nor $[i, 1]$, then the above lemma shows that the only automorphisms of Λ are multiplication by ± 1. Since we have already shown that $\wp$ must take each value of $\mathbb{C}$ exactly twice with multiplicity and that $\wp(z_1) = \wp(z_2)$ iff $z_1 \equiv \pm z_2$ $(mod\ \Lambda)$, the lemma implies the proposition immediately in this case, so we are left with the two special situations.

First, assume Λ is a multiple of $[i, 1]$. By the invariance of the Weber function under lattice isomorphisms $\Lambda \to c\Lambda$ $(c \in \mathbb{C})$, we can assume that

$\Lambda = [i, 1]$. Put $\wp = \wp(\cdot, \Lambda)$. By the above lemma, the automorphism group is cyclic of order four, generated by the map $\sigma : x \mapsto ix$. We must prove that $\wp^2(z_1) = \wp^2(z_2)$ iff $z_1 \equiv \sigma^a(z_2)$ $(mod\, \Lambda)$ for some a. Now $\wp^2(z_1) = \wp^2(z_2)$ iff $\wp(z_1) = \pm\wp(z_2)$. We have

$$(2.1) \qquad \wp(iz) = \wp(iz, \Lambda) = \wp(iz, i\Lambda) = (i)^{-2}\wp(z, \Lambda) = -\wp(z),$$

and the facts that $\wp(z) = \wp(-z)$ and that $\wp$ assumes every complex value exactly twice with multiplicity prove what we want except possibly at points where $z_2 \equiv -z_2$ (since $iz_2 \equiv -iz_2$ iff $z_2 \equiv -z_2$), that is at 2-division points. The non-zero 2-division points of this lattice are $\{\frac{1}{2}, \frac{i}{2}, \frac{1+i}{2}\}$, and we have $\sigma(\frac{1}{2}) = \frac{i}{2}$; $\sigma(\frac{1+i}{2}) = \frac{1+i}{2}$. (2.1) shows that the latter equation implies $\wp(\frac{1+i}{2}) = 0$, and since the values $\wp(\frac{1}{2})$ and $\wp(\frac{i}{2})$ are both assumed with multiplicity two, the proof of this case is finished.

Now we assume $\Lambda = [\rho, 1]$. Then the automorphism group is cyclic of order six, generated by $\sigma : x \mapsto -\rho x$. We must prove that $\wp^3(z_1) = \wp^3(z_2)$ iff $z_1 \equiv \sigma^a(z_2)$ $(mod\Lambda)$ for some a. $\wp^3(z_1) = \wp^3(z_2)$ exactly when $\wp(z_1) = \rho^\epsilon\wp(z_2)$ for $\epsilon = 1$ *or* 2, and the equality

$$(2.2) \quad \wp(\rho^\epsilon z) = \wp(\rho^\epsilon z;\ \Lambda) = \wp(\rho^\epsilon z;\ \rho^\epsilon\Lambda) = \rho^{-2\epsilon}\wp(z;\ \Lambda) = \rho^\epsilon\wp(z;\ \Lambda)$$

together with the 2-to-1-with-multiplicity property of $\wp$ proves what we want except in cases where congruences modulo Λ exist amongst elements of the set $\{\pm z_2, \pm\rho z_2, \pm\rho^2 z_2\}$. Again, calculation shows that this circumstance can only arise when z_2 is a 3-division point, that is if $z_2 \in \{\frac{1}{3}, \frac{2}{3}, \frac{\rho}{3}, \frac{2\rho}{3}, \frac{1+\rho}{3}, \frac{2+2\rho}{3}, \frac{1+2\rho}{3}, \frac{2+\rho}{3}\}$, and that the automorphism group permutes the two sets $\{\frac{1}{3}, \frac{2}{3}, \frac{\rho}{3}, \frac{2\rho}{3}, \frac{1+\rho}{3}, \frac{2+2\rho}{3}\}$ and $\{\frac{1+2\rho}{3}, \frac{2+\rho}{3}\}$ transitively, the element σ^2 fixing the elements of the latter set. (2.2) then tells us that $\wp(\frac{1+2\rho}{3}) = \wp(\frac{2+\rho}{3}) = 0$, and in any case the proof is complete.

We will usually think of the Weber functions $f_0(z, \Lambda)$ as being defined on the elliptic curve $\mathbb{C}/\Lambda$. If the function $t \overset{\phi}{\mapsto} (\wp, \wp')$ $(\wp = \wp(z, \Lambda))$ maps the lattice $\mathbb{C}/\Lambda$ to the elliptic curve E given by the Weierstrass equation

$y^2 = 4x^3 - g_2x - g_3$ then we can define a Weber function on E by

$$h((x, y)) = \frac{g_2g_3}{\Delta}x,$$

if $g_2 \neq 0 \neq g_3$ and the corresponding functions if g_2 or g_3 are zero, and the fact that the function f_0 agrees on two points exactly when they are related by an automorphism of Λ implies that h agrees on two points of E exactly when they are related by an automorphism of E. Furthermore, the function h is clearly defined over the same field that the curve E is defined over.

An immediate corollary of the above theorem, along with the fact that any elliptic curve is isomorphic to $\mathbb{C}/\Lambda$ for some Λ is the result that the automorphism group of any elliptic curve over $\mathbb{C}$ is cyclic of order 2, 4, or 6. In fact, this is a special case of the following theorem.

Theorem 7 *If E is an elliptic curve over a field of characteristic $\neq 2, 3$ then the automorphism group of E is cyclic of degree dividing 6.*

Proof: Basically, an elliptic curve defined over a field of characteristic $\neq 2, 3$ can be assumed given by a Weierstrass equation of the form

$$y^2 = x^3 + ax + b,$$

and one then proves that only changes of variables of the form $x \mapsto u^2x'$, $y \mapsto u^3y'$ for $u \in \bar{K}^*$ both fix the point at infinity (that is are group homomorphisms) and preserve the given equation. Thus any automorphism ϕ of E must be of this form, and since the given map takes the given equation to $u^6(y')^2 = u^6(x')^3 + au^2x' + b$, ϕ being an automorphism implies that $au^{-4} = a$, $bu^{-6} = b$. Hence if $a, b \neq 0$ $u = \pm 1$, while if $a = 0$ (so $b \neq 0$) u must be a fourth root of unity, and if $b = 0$ (so $a \neq 0$) u must be a sixth root of unity. Hence E is cyclic of order 2 4 or 6 depending as whether $ab \neq 0$, $a = 0$ or $b = 0$ For details involving the various changes of variables, see [27, Chapter III, Section 1].

2.4 Special Functions

There are two other specific examples of higher level modular functions which will be important to us, namely the cube root of j and the square root of $j-1728$. We will prove that these functions are weight zero modular of levels three and two respectively. In order to do this, we introduce the ***Dedekind η-function***, defined as follows:

$$\eta(z) = (2\pi)^{\frac{1}{2}} q^{\frac{1}{24}} \prod_{n=1}^{\infty} (1-q^n) = (\Delta(z))^{\frac{1}{24}}$$

Proposition 1 *$\eta(z+1) = e^{\frac{2\pi i}{12}}\eta(z)$, and $\eta(\frac{-1}{z}) = \sqrt{-\imath z}\eta(z)$.*

Proof: The first relation follows immediately from the q-expansion of η. For the second, we know that

$$\Delta(\frac{-1}{z}) = \Delta\left(\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} z\right) = z^{12}\Delta(z);$$

and hence $\left|\eta(\frac{-1}{z})\right| = |\sqrt{z}\eta(z)|$. Since (the principle branch of) $\sqrt{z}$ is holomorphic and non-vanishing on H , as is η, the function

$$\frac{\eta(\frac{-1}{z})}{\sqrt{z}\eta(z)}$$

is holomorphic on H , and has constant modulus 1. Hence by the maximum modulus principle, it is constant, and since $1/i = -i$, evaluating at $z = i$ shows that the constant must be $1/\sqrt{\imath} = \sqrt{-\imath}$, which gives the second relation.

Now $1728 = 2^6 3^3$ and $j = 1728 g_2^3/\Delta$, so

$$\sqrt{j-1728} = \frac{2^3 3^3 g_3}{\eta^{12}} \tag{2.1}$$

$$j^{\frac{1}{3}} = \frac{2^2 3 g_2}{\eta^8}. \tag{2.2}$$

Put $f = \sqrt{\jmath - 1728}$. Then the above proposition along with the fact that g_3 is a modular function of weight 6 tells us that f satisfies $f(z+1) = -f(z)$,

$f(\frac{-1}{z}) = -f(z)$, or to say the same thing that $f(Tz) = -f(z)$ and $f(Sz) = -f(z)$ where S and T are the generators of Γ given above. Consequently $f(T^2z) = f(z)$, since $T^2(z) = z+2$; and if we put $S_2 = TSTTST = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$,

$$f(S_2z) = f(TSTTSTz) = -f(TSTTSz) = -f(Tz) = f(z).$$

Thus to show f is invariant under $\Gamma(2)$, it is sufficient to show the following

Proposition 2 *S_2 and T^2 generate $\Gamma(2)$.*

Proof: Since both S_2 and T^2 are in $\Gamma(2)$, the subgroup generated by them, $<S_2, T^2>$ is contained in $\Gamma(2)$. Given $\gamma_0 \in \Gamma(2)$, pick $\gamma \in <S_2, T^2>$ such that

$$\gamma\gamma_0 = \begin{pmatrix} a & 2b \\ 2c & d \end{pmatrix}$$

has minimal $|c|$. If $c = 0$, then we must have $a = d = \pm 1$, so that $\gamma\gamma_0$ and hence γ_0 is in $<S_2, T^2>$. If $c \neq 0$, pick γ so that not only is $|c|$ minimal, but $|a|$ is minimal among products $\gamma\gamma_0$ with bottom left entry 3c. The fact that

$$\begin{pmatrix} 1 & \pm 2n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 2b \\ c & 2d \end{pmatrix} = \begin{pmatrix} a + 4nc & * \\ 2c & * \end{pmatrix}$$

shows that $a < 2c$, by minimality of a and the fact that $\mathbb{Z}$ is a Euclidean domain. On the other hand, if $a \neq 0$, the fact that

$$\begin{pmatrix} 1 & 0 \\ 2n & 1 \end{pmatrix} \begin{pmatrix} a & 2b \\ c & 2d \end{pmatrix} = \begin{pmatrix} a & * \\ 2c \pm 2na & * \end{pmatrix}$$

shows that $2c < a$, by minimality of c. Hence $a = 0$, but this too is a contradiction since it implies that $|det(\gamma\gamma_0)| \geq 4$. Thus the assumption that the minimal $c \neq 0$ leads us to a contradiction, and we must conclude that $<S_2, T^2> = \Gamma(2)$.

We now turn to $j^{1/3}$. Denote this function by g. Then the above transformation formula for η and the fact that g_2 is a modular function of weight 4 imply that $g\left(\frac{-1}{z}\right) = g(z)$, and $g(z+1) = \rho g(z)$ where ρ is as before a primitive third root of 1. Consequently g is invariant under all matrices in Γ which can be expressed as a word in S and T such the sum of all powers of T appearing is divisible by 3. We show that g is invariant under $\Gamma(3)$ by showing that $\Gamma(3)$ is generated by matrices of this type.

Proposition 3 $\Gamma(3) = <T^3, ST^{-3}S, A, B>$, *where*

$$A = TST^{-3}ST^{-1} = \begin{pmatrix} 4 & -3 \\ 3 & -2 \end{pmatrix} \qquad B = T^2ST^3ST = \begin{pmatrix} -5 & -3 \\ -3 & -2 \end{pmatrix}.$$

Proof: Call the group generated by these elements G. Then $G \subseteq \Gamma(3)$. Conversely, given $\gamma \in \Gamma(3)$, pick $\gamma_0 \in G$ such that $\gamma_0\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has minimal c. If $c = 0$, $a = d = 1$ since $\gamma_0\gamma \in \Gamma(3)$, so $\gamma_0\gamma$ is a power of T^3. If $c \neq 0$, then $a \neq 0$ (else $\gamma_0\gamma$ can't have determinant 1). Since $I = -I$, we can assume that $c > 0$. Multiplying $\gamma_0\gamma$ on the left by T^3 and ST^3S ($= \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix}$) shows that $|a| \leq \frac{3c}{2}$ and $c \leq \frac{3|a|}{2}$. If $a > 0$, $A\gamma_0\gamma = \begin{pmatrix} * & -3a+2c \\ & * \end{pmatrix}$ shows that $|(-3a+2c)| = -3a+2c \geq c$ and hence that $c \geq 3a$, contradicting the above inequality. On the other hand, if $a < 0$, left multiplying $\Gamma_0\gamma$ by B gives the same contradiction, so that in either case we must conclude that $c = 0$ and hence that $\gamma \in \Gamma(3)$.

Now to show that f and g are in F_2 and F_3 respectively, it remains to be shown that they are meromorphic at infinity. Since both f and g have the form $\frac{1}{q}(I)$ where I is a power series in q with constant term 1 and all coefficients in $\mathbb{Z}$, this will be an immediate consequence of the following

Lemma 1 *If $F = 1 + \sum_{n=1}^{\infty} a_n q^n$ is such that $a_n \in \mathbb{Q}$ for all n, then for any m there exists a unique series $1 + \sum_{n=1}^{\infty} b_n q^n$ with $b_n \in \mathbb{Q}$ whose m^{th} power is F.*

Proof: We simply define the b_n inductively, starting with $mb_1 = a_1$. At each n, the equation defining b_{n+1} will be a linear equation in the previous b_i and a_{n+1} with coefficients in $\mathbb{Z}$.

2.5 Ramification in Modular Function Fields

The results of the previous section imply that F_N is ramified over $\mathbb{Q}(j)$ at the ideals (j) and $(j-1728)$ if $N|6$. We now use this fact to gain information about the ramification of F_N over $\mathbb{Q}(j)$ as follows.

Theorem 8 *Let $\mathcal{O}$ be the integral closure of $\mathbb{Q}[j]$ in F_N, $N > 1$. Any ideal $\mathcal{M}$ above $(j - 1728)$ is ramified with index 2. Let z be equivalent to i under Γ, and let M be the maximal ideal equal to the kernel of the map $\mathcal{O} \to \mathbb{C}$ given by $f \mapsto f(z)$. Then $\sigma \in Gal(F_N/\mathbb{Q}(j))$ has $\sigma f_{u,v}(z) = f_{u,v}(z)$ for all second Fricke functions of level N iff $\sigma \in I$, the inertia group of M.*

Proof: (M is above $(j - 1728)$ since $j(z) = j(i) = 1728$.) To begin, assume that N is even and greater than two. We also assume that $z = i$ since $z = Bi$ for some $B \in \Gamma$, and since B permutes second Fricke functions of level N, we will have $\sigma f_{u,v}(z) = f_{u,v}(z)$ if and only if $\sigma f_{u,v}(i) = f_{u,v}(i)$. Since $Gal(F_N/F_1) \cong GL_2(\mathbb{Z}/N\mathbb{Z})/\{\pm I\}$, the action of σ can be represented by a matrix

$$\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathbb{Z}),$$

that is we have $f_{u,v} \mapsto f_{(u,v)\sigma}$. The condition $\sigma f_{u,v}(z) = f_{u,v}(z)$ has now been translated into $f_{(u,v)\sigma}(i) = f_{u,v}(i)$, and this in turn is equivalent to

$$\wp^2\left(\left(\frac{u}{N}, \frac{v}{N}\right)\begin{pmatrix} i & 1 \end{pmatrix}\right) = \wp^2\left(\left(\frac{u}{N}, \frac{v}{N}\right)\sigma\begin{pmatrix} i & 1 \end{pmatrix}\right),$$

where $\wp = \wp(\cdot, [i, 1])$. Now taking $u = 1$, $v = 0$ and $u = 0$, $v = 1$ shows that $\sigma = \pm Id$ or

$$\sigma = \begin{pmatrix} 0 & \pm 1 \\ \mp 1 & 0 \end{pmatrix}, \tag{2.1}$$

since we have already shown that $\wp^2(z_1) = \wp^2(z_2)$ iff $z_1 \equiv \pm z_2$ or $z_1 \equiv \pm i z_2$. $\pm Id$ acts trivially on F_N, and since we know that the inertia group is nontrivial, since $\sqrt{j - 1728} \in F_N$, we conclude that the inertia group must be exactly the group of order two generated by the matrix (2.1) above. Now the same result holds if N is odd or $N = 2$, since in this case we have the conclusion for F_{2N}, and the restriction of $\sigma \in I$ from F_{2N} to F_N is represented by the same matrix on F_N as on F_{2N}, and must act non-trivially since it sends any Fricke function of the first kind to its negative. Hence the conclusion holds for any N.

Now by applying the same technique to an ideal over (j) we obtain the following

Theorem 9 *Let $\mathcal{O}$ be the integral closure of $\mathbb{Q}[j]$ in F_N, $N > 1$. Any ideal $\mathcal{M}$ above (j) is ramified with index 3. Let z be equivalent to ρ under Γ, and let M be the maximal ideal equal to the kernel of the map $\mathcal{O} \to \mathbb{C}$ given by $f \mapsto f(z)$. Then $\sigma \in Gal(F_N/\mathbb{Q}(j))$ has $\sigma f_{u,v}(z) = f_{u,v}(z)$ for all third Fricke functions of level N iff $\sigma \in I$.*

Furthermore, since the non-triviality of the inertia groups in the above 'special' cases come directly from the use of the second and third Fricke functions which involve $\wp^2$ and $\wp^3$ respectively, we can also conclude by using the same technique that if z is not equivalent to i or ρ under Γ, $\sigma f_{u,v}(z) = f_{u,v}(z)$ for all u, v iff $\sigma = Id$ - that is that the ideals $\mathcal{M}$ of $\mathcal{O}$ not above (j) or$(j - 1728)$ are unramified in the extension $F_N/\mathbb{Q}(j)$ (when $\mathbb{Q}[j]$ is the underlying domain).

2.6 Results from Algebraic Number Theory

2.6.1 ζ and L Functions, and Density

Throughout this section, the phrase 'prime ideal' means 'nonzero prime ideal'.

Definition: If K is a number field (finite extension of $\mathbb{Q}$) and $\mathcal{O}$ its ring of integers, the *Dedekind zeta function for K* is defined for $\Re(z) > 1$ by

$$\zeta_K(z) = \sum_{\mathfrak{a} \lhd \mathcal{O}} \frac{1}{N(\mathfrak{a})^z},$$

the sum ranging over all nonzero ideals $\mathfrak{a}$ of $\mathcal{O}$, and with the norm $N(\mathfrak{a})$ the norm from K to $\mathbb{Q}$ of $\mathfrak{a}$, given by $N(\mathfrak{a}) = |\mathcal{O}/\mathfrak{a}|$.

This function is analytic in $\Re(z) > 1$ and $(z-1)\zeta_K(z)$ can be extended to an entire function such that $\zeta_K(z)$ has a simple pole at $z = 1$. Also,

$$\log \zeta_k(z) = \sum_{m=1}^{\infty} \sum_{\mathfrak{p}} \frac{1}{mN(\mathfrak{p})^{mz}},$$

with the inner sum running over all non-zero prime ideals of $\mathcal{O}$. The part of the sum corresponding to $m \geq 2$ is analytic in $\Re(z) > 1/2$, so only the $m = 1$ sum has significance for evaluating the behaviour of $\log \zeta_k$ near $z = 1$. Thus, the fact that ζ_k has a simple pole at 1 implies that

$$\log \zeta_k(z) \sim \sum_{\mathfrak{p}} \frac{1}{N(\mathfrak{p})^z} \sim \log\left(\frac{1}{z-1}\right),$$

where $f \sim g$ means that $f - g$ is analytic in a neighbourhood of 1. This leads us to the following

Definition: Let S be a set of prime ideals of K. We define the *(Dirichlet) density of S*, to be the limit

$$\lim_{z \to 1} \left(\log \frac{1}{z-1}\right)^{-1} \left(\sum_{\mathfrak{p} \in S} \frac{1}{N(\mathfrak{p})^z}\right)$$

if it exists. If R and S are two sets of primes of K, we define $R \prec S$ to mean that $R \setminus S$ has density 0.

Note that all finite sets of primes have density zero, and that if S_K is the set of primes of $\mathbb{Z}$ which split completely in $\mathcal{O}$, that the density of the complement of S_K is zero, since any prime which is not above a completely split prime has norm at least a square of a prime of $\mathbb{Z}$, and since there are at most $[K : \mathbb{Q}]$ primes of $\mathcal{O}$ over a given prime of $\mathbb{Z}$, the sum over all these primes converges near one. Thus, if S is any set of primes of K, the density of S is equal to the density of $S \bigcap S_K$, (in the sense that if one of these densities exists, so does the other and they are equal) since $S \setminus S \bigcap S_K$ has density zero.

Now given an extension K/k of number fields, define $S_{K/k}$ to be the set of primes of k which split completely in K.

Proposition 1 *If K/k is Galois, the density of $S_{K/k}$ is equal to $1/[K : k]$.*

Proof: The density of $S_{K/k}$ is

$$\lim_{z\to 1}\left(\log\frac{1}{z-1}\right)^{-1}\left(\sum_{\mathfrak{p}\in S_{K/k}}\frac{1}{(N\mathfrak{p})^z}\right).$$

By the above comments this is equal to the same expression with the sum extending over only the primes of k which are above primes of $\mathbb{Z}$ splitting completely in k and which split completely in K, that is over the primes of k which lie below primes of K which split completely in $K/\mathbb{Q}$. Since K/k is Galois, there are exactly $[K : k]$ primes of K above any prime of k which splits completely in K, so the density is equal to

$$\lim_{z\to 1}\left(\log\frac{1}{z-1}\right)^{-1}\left(\frac{1}{[K:k]}\sum_{\mathrm{P}\in S_{L/\mathrm{Q}}}\frac{1}{(N\mathfrak{P})^z}\right) = \frac{1}{[K:k]},$$

as required.

Corollary *If $k \subseteq K \subseteq L$ are Galois extensions of number fields and $S_{K/k} \prec S_{L/k}$ then $L = K$.*

Proof: Since $K \subseteq L$, $S_{L/k} \subseteq S_{K/k}$, so the two sets $S_{K/k}$ and $S_{L/k}$ differ by a set of density zero. Hence their densities are equal, and by the proposition this implies that $[L : k] = [K : k]$.

Proposition 2 *If K and L are finite Galois extensions of the number field k then $S_{K/k} \prec S_{L/k}$ if and only if $L \subseteq K$.*

Proof: K/k and L/k both Galois implies that KL/k is Galois, since any injection $KL \hookrightarrow \mathbb{C}$ fixing k restricts to give injections $K \hookrightarrow \mathbb{C}$ and $L \hookrightarrow \mathbb{C}$, which fix k and hence by normality of K/k and L/k are actually maps $K \to K$ and $L \to L$, so $KL \to KL$. If $L \subseteq K$ then $S_{K/k} \subseteq S_{L/k}$, so certainly $S_{K/k} \prec S_{L/k}$. On the other hand, if $S_{K/k} \prec S_{L/k}$, then the equality

$$S_{KL/k} = S_{K/k} \bigcap S_{L/k} \tag{2.1}$$

shows that $S_{K/k} \prec S_{KL/k}$, so that $KL = K$ and thus $L \subseteq K$. To prove (2.1) let $\mathfrak{p}$ be a prime of k that splits completely in K/k and L/k. Since this implies that $\mathfrak{p}$ splits completely in $K/K \bigcap L$ and $L/K \bigcap L$, it is enough to prove that $\mathfrak{p}$ splits completely in KL under the assumption that K and L are linearly disjoint. Assuming this, we have a group isomorphism

$$Gal(K/k) \times Gal(L/k) \tilde{\to} Gal(KL/k).$$

If D_K and D_L are the decomposition groups of (fixed) primes over $\mathfrak{p}$ in K and L respectively, then $|D_K| = |D_L| = 1$ since $\mathfrak{p}$ splits completely in these fields. But the isomorphism sends $D_K \times D_L$ to the decomposition group of a prime of KL above $\mathfrak{p}$, so $\mathfrak{p}$ splits completely in KL. The other containment of (2.1) follows from the $efg = n$ equality for Galois extensions of number fields.

Corollary *If K and L are finite Galois extensions of the field k and $S_{K/k}$ differs from $S_{L/k}$ by a set of density zero, then $L = K$.*

Proof: This is immediate, since the hypothesis implies both $S_{K/k} \prec S_{L/k}$ and $S_{L/k} \prec S_{K/k}$.

In the study of quadratic fields to follow we will make use of the Dirichlet L-function in addition to the Dedekind ζ-function defined above. If $-d < 0$ is the discriminant of an imaginary quadratic field, then there exists a primitive real character $\chi : \mathbb{Z}/d\mathbb{Z} \to \{\pm 1\}$ (see [5]). For this character, we define the *Dirichlet L-function* to be

$$L(s) = L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}.$$

This series converges absolutely for $\Re(s) > 1$, and in fact can be analytically continued to an entire function. Its value at $s = 1$ encodes information about the class number and other invariants of the field $\mathbb{Q}(\sqrt{-d})$.

2.6.2 Results from Class Field Theory

Class field theory is the study of Abelian extensions of number fields - that is, Galois extensions of number fields whose Galois groups are Abelian. The main theorems in this subject provide a classification of the finite abelian extensions of a given number field in terms of subgroups of group of fractional ideals of the (integer ring of) the number field. These theorems were developed and proved during the end of the nineteenth century and the early twentieth century by the combined efforts of many mathematicians, beginning with Kronecker and Weber, and including Hilbert and E. Artin. We will provide here a cursory summary of the main theorems of the theory, and to avoid having to introduce many new notions and notations we will stick to the most basic form of the theory. Class field theory has its beginning in the Frobenius automorphism from algebraic number theory. If k is a number field with ring of integers $\mathfrak{o}$ and K is a finite extension of k with ring of integers $\mathcal{O}$, then for any prime $\mathfrak{P} \lhd \mathcal{O}$ above $\mathfrak{p} \lhd \mathfrak{o}$, there is an associated extension of residue fields $\mathfrak{o}/\mathfrak{p} \hookrightarrow \mathcal{O}/\mathfrak{P}$. Since all ideals of an integer ring are finite indexed, this is an extension of finite fields, and consequently a Galois extension with cyclic Galois group, generated by the element $\tilde{\sigma}$ which takes $\tilde{x} \in K/\mathfrak{P}$ to x^{p^f}, where $p^f = |\mathfrak{o}/\mathfrak{p}|$. This element is called the *Frobenius automorphism* of $K/\mathfrak{P}$, and any element $\sigma \in Gal(K/k)$ which reduces to the automorphism $\tilde{\sigma}$ on $K/\mathfrak{P}$ is called a *Frobenius element for* $\mathfrak{P}$. These elements are characterized by the fact that they satisfy the equation

$$\sigma(x) \equiv x^{p^f} \ (mod\ \mathfrak{P})$$

for all $x \in K$. If the prime $\mathfrak{p}$ is unramified in the extension K/k, then (by the triviality of the inertia group $E(\mathfrak{P}/\mathfrak{p})$) the element σ is unique, and if $\mathfrak{Q} = \tau(\mathfrak{P})$ ($\tau \in Gal(K/k)$) is another prime of $\mathcal{O}$ over $\mathfrak{p}$, the Frobenius element for $\mathfrak{Q}/\mathfrak{p}$ is conjugate to the Frobenius element for $\mathfrak{P}/\mathfrak{p}$, as is easily seen by examining the equation characterizing this element. Thus, if the extension K/k is Abelian, all Frobenius elements for primes over $\mathfrak{p}$ are equal. Assume from now on that the extension K/k is Abelian. Let S denote the set of prime ideals of k which are ramified in the extension K/k, and let I_S denote the group of fractional $\mathcal{O}$-ideals which are relatively prime to all primes in S. The above discussion shows that for any prime of $\mathcal{O}$ which is in I_S, there exists a unique Frobenius element of $Gal(K/k)$. This assignment can be extended multiplicatively to yield a homomorphism

$$\phi : I_S \to Gal(K/k).$$

This map is called the *Artin map*. Notice that the kernel of the Artin map contains the subgroup of I_S generated by the set of primes $\mathfrak{p}$ which split completely in K/k, since a prime splits completely exactly when the residue extension $k/\mathfrak{p} \hookrightarrow K/\mathfrak{P}$ is trivial for some (every) prime $\mathfrak{P}$ over $\mathfrak{p}$, and this will happen exactly when the Frobenius automorphism is trivial. Thus, to each finite abelian extension K/k unramified outside of S there corresponds a subgroup of I_S, the kernel of the Artin map. We will call an Abelian extension unramified outside of S an S-extension. Proposition 2 of the previous section allows us to give a quick proof of

Proposition 3 *The map ϕ above is a surjection.*

Proof: Let L be the fixed field of $\phi(I_S)$. Since K/k is Galois, it is enough to show $L = k$. But for all primes $\mathfrak{p} \in I_S$, $\phi(\mathfrak{p})$ is trivial on L - that is, $\mathfrak{p}$ splits completely in L. Thus, all but finitely many primes of k split completely in L, so $L = k$.

One of the principal achievements of class field theory is to answer the question of which subgroups of I_S can occur as kernels of Artin maps (Artin kernels) for Abelian extensions of k.

Definition: If $\mathfrak{a}$ is an ideal of $\mathcal{O}$ and $x \in \mathfrak{o}$, we define $x \equiv 1$ $(\mathit{mod}^*\ \mathfrak{a})$ to mean that

- For all $\sigma \in Gal(k/\mathbb{Q})$ such that $\sigma(k) \subseteq \mathbb{R}$, $\sigma(x) > 0$ (x is totally positive), and
- If $\mathfrak{p}^n | \mathfrak{a}$, $x \equiv 1\ (\mathit{mod}\ \mathfrak{p}^n)$.

Let S be the set of primes dividing $\mathfrak{a}$. Then the *principal* $\mathfrak{a}$*-subgroup* of the ideal group I_S is

$$P_{\mathrm{a}} = \{\mathfrak{a} \in I :\ \mathfrak{a} = (\alpha/\beta) \text{ for some } \alpha,\ \beta \in \mathfrak{o}, \alpha \equiv \beta \equiv 1\ (\mathit{mod}^*\ S)\}$$

Using this definition, the result we require from class field theory states that

Theorem 10 *With notations as above, a subgroup* $H \leq I_S$ *occurs as the Artin kernel for some* S*-ramified Abelian extension* K/k *if and only if* $P_{\mathrm{a}} \leq H$ *for some ideal* $\mathfrak{a} \lhd \mathcal{O}$ *divisible only by primes in* S*.*

Thus, in particular, there exists an Abelian extension K whose Artin kernel is *exactly* P_{a}. This particular extension is called the *ray class field of k to the modulus* $\mathfrak{a}$. By the definition of the Artin kernel, the primes of k which split completely in this particular extension are exactly those which are principal, and can be generated by an element $\alpha \equiv 1\ (\mathit{mod}^*\ \mathfrak{a})$.

Notice that in the case where k is an imaginary quadratic field, the totally positive condition in the definition of mod^* is vacuous, and hence $x \equiv 1\ (\mathit{mod}^* \mathfrak{a})$ if and only if $x \equiv 1\ (\mathit{mod}\ \mathfrak{a})$. In this case, if we take S to be the empty set, the theorem shows that any Abelian unramified extension of k must have Artin kernel contained in the set of principal ideals - and, since the relation between fields and Artin kernels is (strictly) inclusion reversing, we have that the maximal unramified Abelian extension of an imaginary quadratic field k is characterized by the fact that the primes which split in it are exactly the principal primes of k. The maximal everywhere unramified Abelian extension of a number field k is called the *Hilbert class field of k*.

We end this section by proving the famous Kronecker-Weber theorem, assuming theorem 10 above.

Theorem 11 (Kronecker-Weber) *If k is an Abelian extension of $\mathbb{Q}$, then there exists some integer m such that $k \leq \mathbb{Q}(\zeta_m)$, where ζ_m is a primitive m^{th} root of unity.*

Proof: Let $\mathfrak{o}$ be the integer ring of k, and let S be the set of prime ideals of $\mathbb{Z}$ which ramify in $\mathfrak{o}$. By theorem 10 there exists some ideal $(m) \triangleleft \mathbb{Z}$ such that $P_{(m)} \leq H$, where H is the Artin kernel of k. Thus, all primes of $\mathbb{Z}$ which are congruent to one modulo m split completely in H. Since the primes which split in the extension $\mathbb{Q}(\zeta_m)/\mathbb{Q}$ are exactly those congruent to one modulo m, we conclude that $S_{L/\mathbb{Q}} \subseteq S_{k/\mathbb{Q}}$, and since both extensions are Galois, the result follows by proposition 2 of this section.

Chapter 3

abc Implies No 'Siegel Zeroes'

The *abc*-conjecture, while quite obviously a powerful tool in the study of diophantine equations, is perhaps not so obviously suited for application to situations of a more analytic nature. For this reason, the recent result of Granville and Stark [10] that the uniform *abc*-conjecture implies the non-existence of so-called Siegel zeroes of Dirichlet L-functions 'attached' to quadratic fields is quite interesting. If $-d < 0$ is the discriminant of an imaginary quadratic field (so that $-d$ is either square free and congruent to one modulo four, or $d = 4d'$ with d' squarefree and congruent to two or three modulo four), then there exists a real primitive Dirichlet character to the modulus d, and furthermore this character encodes information about the splitting of primes of $\mathbb{Z}$ in the extension $\mathbb{Z}[\sqrt{-d}]$. Of course, there is a generalized Riemann hypothesis for these functions which states that they have no zeroes on the 'critical strip' $\{z \in \mathbb{C} : \ 0 \leq \Re(z) \leq 1\}$ except on the line $\Re(z) = 1/2$; but this problem is so difficult that even the much weaker problem of showing that the functions have no zeroes in $I_{c,d} = \{x \in \mathbb{R} : \ 1 - c/\log(d) \leq x \leq 1\}$ for some positive constant c is unsolved. It has been shown (see Davenport [5]) that for a given $-d$, there is at most one simple zero in this interval, and the existence of such a 'Siegel zero' would have ramifications for other problems in number theory, including asymptotic estimates for the size of least quadratic non-residues mod p and various sieve estimates.

Granville and Stark's result rests on a 1934 result of Mahler [18], who showed that the vanishing of the L-functions of imaginary quadratic fields on the interval mentioned above has implications for the class numbers of imaginary quadratic fields. Specifically, assuming a certain bound on the class numbers as $-d$ grows, Mahler proved that there exists a contant $c > 0$ so that no L-function $L(x, \chi)$ has a zero on $I_{c,d}$. Granville and Stark's paper proves that the requisite bound on the class numbers is implied by the uniform *abc*-conjecture by studying solutions of a certain diophantine equation which arises through the theory of modular functions. Having outlined general modular function theory in the previous chapter, we now are in a postion to give a complete exposition of the work which directly underlies Granville and Stark's result.

3.1 The Modular Equation

The purpose of this section is to show that the function j takes on algebraic integer values at imaginary quadratic arguments. Later, it will be shown that in fact the j-function evaluated at appropriate arguments generates various abelian extensions of imaginary quadratic fields.

Definition: $H_n = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathbb{Z}) : ad - bc = n, \gcd(a,b,c,d) = 1\}$.

If $G = \mathbb{Z}x \bigoplus \mathbb{Z}y$ is a free abelian group of rank 2, then for any

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in H_n,$$

$H = \alpha(G) = \mathbb{Z}(ax + by) \bigoplus \mathbb{Z}(cx + dy)$ is a subgroup of G of index n, and the elementary divisor theorem says that there exists a basis $\{x', y'\}$ of G such that $H = \mathbb{Z}e_1x' + \mathbb{Z}e_2y'$, with $e_1|e_2$. e_1 is then an integer which divides each of a, b, c, d and hence $e_1 = 1$. Thus, there exists an element $\gamma \in \Gamma$ such that $\gamma\alpha\gamma^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix}$. Hence, the coset of $\begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix}$ is sent to each coset $\Gamma\alpha$ of the left action of Γ on H_n by the right action of Γ, and hence Γ operates right transitively on the left cosets $\Gamma\alpha$ of $\Gamma \setminus H_n$. Furthermore, a calculation shows that

$$\left\{\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} : ad = n, \quad a > 0, \quad 0 \leq b < n\right\}$$

represent a complete set of distinct coset representatives of $\Gamma \setminus H_n$. We will denote the number of these matrices by $\psi(n)$, and the matrices themselves by $\{\alpha_i\}_{i=1}^{\psi(n)}$.

Definition: The *modular polynomial of level n* is the polynomial

$$\Phi_n(X, j) = \prod_{i=1}^{\psi(n)} (X - j \circ \alpha_i).$$

The coefficients of Φ_n (viewed as a polynomial in X) are elementary symmetric functions in the $j \circ \alpha_i$. They are holomorphic on $\mathcal{H}$ since all the functions $\alpha_i = \frac{az+b}{d}$ are holomorphic maps $\mathcal{H} \to \mathcal{H}$, and j is holomorphic on $\mathcal{H}$.

Proposition 1 *The coefficients of Φ_n are meromorphic at ∞, and are polynomials in j with integer coefficients, that is $\Phi_n(X,j) \in \mathbb{Z}[X, j]$.*

Proof: $j \circ \alpha_i \circ \gamma(z) = j \circ \gamma' \alpha_l(z) = j\alpha_l(z)$ (where $\alpha_i\gamma = \alpha_l\gamma'$ with $\gamma' \in \Gamma$), since the $\{\alpha_i\}$ are coset representatives of $\Gamma \setminus H_n$ and by Γ-invariance of j. Thus Γ permutes the $j \circ \alpha_i$, so the coefficients of Φ_n are invariant under Γ. Hence if f is any one of these coefficients, f defines a holomorphic function on $\mathbb{D} \setminus \{0\}$, and hence f has a Laurent expansion (in q) at zero, so the meromorphicity of f at zero is equivalent to its being bounded in absolute value by some power of $1/q$ near zero. This clearly holds for f, since each of the $j \circ \alpha_i$ are bounded by a power of $1/q$ near zero. Thus each f is a polynomial in j with coefficients in $\mathbb{Z}(\zeta_n)$ where ζ_n is any primitive n^{th} root of unity. (Since each function $j \circ \alpha_i$ has q-series coefficients in $\mathbb{Z}(\zeta_n)$). Furthermore each $j \circ \alpha_i$ is in the field $\mathbb{Q}(\zeta_n)((q^{1/n}))$, and is invariant under $\mathbb{Q}((q^{1/n}))$-automorphisms of that field, thus is in $\mathbb{Z}((q^{1/n}))$.

Theorem 12 *(a) $\Phi_n(X, j)$ is irreducible over $\mathbb{C}(j)$. (b) $\Phi_n(X, j) = \Phi_n(j, X)$. (c) If n is not a perfect square, $\Phi_n(j, j)$ is a polynomial in j of degree > 1 with leading coefficient ± 1.*

Proof: (a) is true since $\Gamma/\Gamma(n) = Gal(F_n/C(j))$, and Γ permutes the elements $j \circ \alpha_i$, that is the roots of Φ_n, transitively.

For (b), note first that since the matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \text{ and } \begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix}$$

are both among the α_i, we have both $\Phi_n(j(z), j(nz)) = 0$ and $\Phi_n(j(z), j(z/n)) = 0$, and on making the substitution $z \mapsto nz$ the second equation becomes $\Phi_n(j(nz), j(z)) = 0$. Hence $\Phi_n(X, j)$ and $\Phi_n(j, X)$ have a common root, namely $j \circ \begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix}$. Since $\Phi_n(X, j)$ is irreducible, this implies that $\Phi_n(X, j)$ divides $\Phi_n(j, X)$, that is $\Phi_n(j, X) = \Phi_n(X, j)A(X, j)$ for some $A \in \mathbb{Z}[X, j]$. But then we must have $\Phi_n(X, j) = \Phi_n(j, X)A(j, X)$, and therefore that $A(X, j)A(j, X) = 1$. Inasmuch as the only units in $\mathbb{Z}[X, j]$ are ± 1, we must have $A(X, j) = \pm 1$. $A(X, j) = -1$ is impossible since this would imply that $\Phi_n(X, j) = -\Phi_n(j, X)$, and hence that $\Phi_n(j, j) = 0$, that is that j is a root of $\Phi_n(X, j)$, contradicting part (a).

Now to prove (c), assume n is not a square. Then $n = ad$ implies $a \neq d$, and so in the q-expansion of $j \circ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$, the leading coefficient is

$$\frac{1}{\zeta_d^b q^{a/d}} \neq \frac{1}{q},$$

so the leading coefficient of each $j - j \circ \alpha_i$ is a root of unity. Hence the leading coefficient in the q-expansion of $\Phi_n(j, j)$ is also a root of unity, but this element is in $\mathbb{Z}[j]$, since $\Phi_n(j, j) \in \mathbb{Z}[j]$, so that the leading coefficient must be ± 1.

This proposition shows that $j \circ \alpha_i$ is integral over $\mathbb{Z}[j]$, since it is a root of $\Phi_n(X, j)$ which is a polynomial in $\mathbb{Z}[X, j]$ with leading coefficient 1. Therefore for any $\alpha \in M_2^+(\mathbb{Q})$, $j \circ \alpha$ is integral over $\mathbb{Z}[j]$, since it has the same action on $\mathcal{H}$ as some integral matrix, and hence the same action as some coset representative for $\Gamma \setminus H_n$ for some n.

Theorem 13 *If $z \in \mathcal{H}$ is imaginary quadratic, then $j(z)$ is an algebraic integer.*

Proof: Say $\mathbb{Q}(z) = \mathbb{Q}(\sqrt{-m})$, with m a squarefree positive integer. Let $\mathcal{O}$ be the ring of integers of $\mathbb{Q}(\sqrt{-m})$, say $\mathcal{O} = [1, \tau]$, so that $\tau = \sqrt{-m}$ or $\tau = \frac{1+\sqrt{-m}}{2}$, according as whether $-m \equiv 2, 3 \pmod 4$ or $-m \equiv 1 \pmod 4$. If $m = 1$, take $\lambda = 1 + \imath$, and if $m > 1$, take $\lambda = \sqrt{-m}$. Then $N(\lambda) = \lambda\bar{\lambda}$ is squarefree, and by the definition of the norm, if $\lambda\tau = a\tau + b$, $\lambda = c\tau + d$ then

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

has determinant $N(\lambda)$. Furthermore $\alpha\tau = \lambda\tau/\lambda = \tau$, so $j(\tau)$ is a root of $\Phi_{N(\lambda)}(X, X)$, and is consequently an algebraic integer. Now since $z \in \mathbb{Q}(\tau)$, $\tau = uz + v$ with $u, v \in \mathbb{Q}$, that is that $\tau = \beta z$ for some $\beta \in M_2^+(\mathbb{Z})$, the determinant of β being positive since $u > 0$, as z and τ are both in $\mathcal{H}$. But by the above comment, $j(\beta\tau)$ is integral over $\mathbb{Z}[j(\tau)]$, and since integrality is transitive, this implies that $j(z)$ is an algebraic integer.

3.2 Generation of Class Fields

3.2.1 Hilbert Class Field

In this section we will show that when k is an imaginary quadratic field, the Hilbert class field of k is generated by numbers of the form $j(z)$ for $z \in k$. A few notions from algebraic geometry are needed.

If C_1 and C_2 are nonsingular projective curves over some algebraically closed field K with function fields $K(C_1)$ and $K(C_2)$ then any algebraic map

$\phi : C_1 \to C_2$ induces a map on function fields $\phi^* : K(C_2) \to K(C_1)$ given by $\phi^*(g) = g \circ \phi$. If ϕ is non constant, then the image $\phi(C_1)$ is a dense subset of C_2 and hence the map ϕ^* is non-zero and hence an injection. In this case, the *degree of* ϕ is defined to be the degree of the extension $K(C_1)/\phi^*(K(C_2))$ (which one can prove to be finite). If the extension is separable, the map ϕ is said to be separable, and this condition is equivalent to the induced map ϕ^* on differentials being non-zero. Corresponding to the Frobenius automorphism in number theory, there is a Frobenius map. If F is a field of characteristic $p > 0$, and C is a (non-singular) curve over F defined by some equation $f(x, y) = 0$, so that the function field of C is $\bar{F}(C) = \bar{F}(X, Y)/(f(X, Y))$ ($\bar{F}$ denoting the algebraic closure of F) then since the map $\alpha \mapsto \alpha^p$ is an automorphism of F, if we define a curve $C^{(p)}$ by the equation $f(x^p, y^p) = 0$ the function field of $C^{(p)}$ is contained in that of C, and the extension of function fields $\bar{F}(C)/\bar{F}(C^{(p)})$ is purely inseparable of degree p. Since the curve C is defined over F, this means that the extension $F(C)/F(C^{(p)})$ is also (purely inseparable) of degree p. The map $C \to C^{(p)}$ given by $(x, y) \mapsto (x^p, y^p)$ is called the *Frobenius map.*

Theorem 14 *Let* $k = \mathbb{Q}(\sqrt{-d})$ *($d \in \mathbb{Z}$, $d > 0$ square free) be an imaginary quadratic field with integer ring* $\mathcal{O} = \mathcal{O}_k$ *and* $a \subseteq \mathcal{O}$ *an ideal. Then the smallest Galois extension of* k *containing* $j(\mathfrak{a})$ *is the Hilbert class field to* k*, and if* $\mathfrak{a}_i$ *($i = 1, 2, \cdots, h$) are representatives of all ideal classes of* k *then* $\{j(\mathfrak{a}_i)\}$ *represent a full set of conjugates of* $j(\mathfrak{a})$ *over* k.

Proof: Let K be the smallest Galois extension of k containing $j(\mathfrak{b})$ for every integral $\mathcal{O}$-ideal $\mathfrak{b}$. Since j takes different values on different lattices (ideals) exactly when they are not scalar multiples of each other (in different ideal classes), this extension is of finite degree. By the results of 2.6.1 and 2.6.2, to prove that K is the Hilbert class field of k it is enough to prove that all but finitely many split primes of k split in K if and only if they are principal, since this will show that the set of primes of k which split completely in K differs from the set of primes k which split completely in the Hilbert class field by a set of density zero. For each i, $1 \leq i \leq h$, pick an elliptic curve E_i defined over K with j-invariant $j(\mathfrak{a}_i)$ and discriminant Δ_i. Then $\mathbb{C}/\mathfrak{a}_i \cong E_i$ (analytic isomorphism), and $End(E_i) \cong \mathcal{O}$. Pick a prime $p \in \mathbb{Z}$ such that $p = \mathfrak{p}\mathfrak{p}'$ is split in k, and $p \notin S$ where

$$S = \{p | \Delta_i \text{ for some } i, \text{ or } p | \prod_{a<b} (\xi_a - \xi_b)\} \bigcup \{2, 3\},$$

with ξ_a all distinct values of $\sigma j(\mathfrak{a}_i)$ for $\sigma \in Gal(K/k)$ and $1 \leq i \leq h$. For later reference, we will call this product D (of course, the theorem implies that D is just the discriminant of the element $j(\mathfrak{b})$ for any integral ideal $\mathfrak{b}$). Since $p \not| \Delta_i$ there exists a prime $\mathfrak{P}$ of K over $\mathfrak{p}$ not dividing Δ. Without loss of generality we can assume that $\mathfrak{a} = \mathfrak{a}_i$ for some i; put $A = E_i$. Then A is given by a Weierstrass equation $y^2 = x^3 + ax + b$, with $a, b \in K$.

Pick a prime b of K such that $N(\mathfrak{b})$ is prime to $p = N(\mathfrak{p})$ and $\mathfrak{p}\mathfrak{b} = (\alpha)$ is principal. Then $(\alpha)\mathfrak{p}^{-1}\mathfrak{a} = \mathfrak{b}\mathfrak{a}$, and we have the following commutative diagram

$$\begin{tikzcd}
\mathbb{C} \arrow[r] \arrow[d] & \mathbb{C} \arrow[r] \arrow[d] & \mathbb{C} \arrow[r] \arrow[d] & \mathbb{C} \arrow[d] \\
\mathbb{C}/\mathfrak{a} \arrow[r] \arrow[d] & \mathbb{C}/\mathfrak{p}^{-1}\mathfrak{a} \arrow[r, "\sim"] \arrow[d] & \mathbb{C}/\mathfrak{b}\mathfrak{a} \arrow[r] \arrow[d] & \mathbb{C}/\mathfrak{a} \arrow[d] \\
A \arrow[r, "\lambda"] & B \arrow[r, "\sim"] & B \arrow[r, "\mu"] & A
\end{tikzcd}$$

with λ and μ some isogenies, and the maps $\mathbb{C}/\mathfrak{a} \to \mathbb{C}/\mathfrak{p}^{-1}\mathfrak{a}$ and $\mathbb{C}/\mathfrak{b}\mathfrak{a} \to \mathbb{C}/\mathfrak{a}$ induced by the inclusions $\mathfrak{a} \hookrightarrow \mathfrak{p}^{-1}\mathfrak{a}$ and $\mathfrak{b}\mathfrak{a} \hookrightarrow \mathfrak{a}$ respectively. The composite of the top row is the multiplication by α map, and the maps $\mathbb{C}/\mathfrak{p}^{-1}\mathfrak{a} \to B$ and $\mathbb{C}/\mathfrak{b}\mathfrak{a} \to B$ are possibly different, so that the target curve is B in both cases. Consequently,

$$pN(\mathfrak{b}) = N(\mathfrak{p}\mathfrak{b}) = \left| Ker \left(\frac{\mathbb{C}}{\mathfrak{a}} \to \frac{\mathbb{C}}{\mathfrak{p}^{-1}\mathfrak{a}} \tilde{\to} \frac{\mathbb{C}}{\mathfrak{b}\mathfrak{a}} \to \frac{\mathbb{C}}{\mathfrak{a}} \right) \right| = [\mathbb{C}(A) : (\mu \circ \ \lambda)^* \mathbb{C}(A)]$$

and the right hand side is equal to the degree of $\mu \circ \lambda$. Since the maps μ and λ are defined over K, as is A, this degree is also equal to the degree of the extension $[K(A) : (\mu \circ \ \lambda)^* K(A)]$. Similarly, $\mu : B_K \to A_K$ has degree $N(\mathfrak{b})$, and $\lambda : A_K \to B_K$ has degree p. Now the map μ reduces to give a map

$$\tilde{\mu} : B_{\tilde{K}} \to A_{\tilde{K}}$$

where $\tilde{K} = K/\mathfrak{P}$, and this map has degree $N(\mathfrak{b})$ - prime to p. Thus the map $\tilde{\mu}$ is separable. But the map $(\mu \circ \lambda)^*$, being the pull back of a map corresponding to multiplication by α, acts as multiplication by α on differentials. Since $\mathfrak{P}|\alpha$, this means that the map $(\tilde{\mu} \circ \tilde{\lambda})^*$ sends all differentials to zero, and hence is inseparable. Thus $\tilde{\lambda}$ is inseparable, but this map has degree (at most) p (being the reduction of a map of degree p). This means that $\tilde{\lambda}$ is purely inseparable of degree p. Consequently the function field of $A_{\tilde{K}}$ is an inseparable extension of the function field of $B_{\tilde{K}}$ of degree p. But the function field of $A_{\tilde{K}}$ is $\tilde{K}(x, y)$, and this field is an inseparable extension of degree p of exactly one field - that is $\tilde{K}(x^p, y^p)$ ($\tilde{K}(x, y)$ is a degree p inseparable extension of $\tilde{K}(x^p, y^p)$, and any other field over which $\tilde{K}(x, y)$ is a degree p inseparable extension must contain x^p and y^p, hence must equal $\tilde{K}(x^p, y^p)$). Thus

$$\tilde{K}(B) \cong \tilde{K}(x^p, y^p),$$

and since nonsingular projective curves are determined up to birational equivalence by their function fields this implies that

$$B_{\tilde{K}} \cong A^p_{\tilde{K}},$$

so the j-invariant of $B_{\tilde{K}}$ must equal $j(A^p_{\tilde{K}}) = j(A)^p$. Since this equality is as elements of $\tilde{K} = K/\mathfrak{P}$, we have that

$$j(\mathfrak{a})^p = j(A)^p \equiv j(B) = j(\mathfrak{p}^{-1}\mathfrak{a}),$$

with the congruence being modulo $\mathfrak{P}$. Letting σ_{P} denote the Frobenius automorphism of K/k associated to $\mathfrak{P}/\mathfrak{p}$, for all $z \in K$ we have $\sigma_{\mathrm{P}}(z) \equiv z^{N(\mathfrak{p})}$ $(mod\, \mathfrak{P})$, and since $N(\mathfrak{p}) = p$ we have $\sigma_{\mathrm{P}}(j(\mathfrak{a})) \equiv j(\mathfrak{p}^{-1}\mathfrak{a})$ $(mod\, \mathfrak{P})$, but this means that $\mathfrak{P}|(\sigma_{\mathrm{P}}(j(\mathfrak{a})) - j(\mathfrak{p}^{-1}\mathfrak{a}))$, so

$$\sigma_{\mathrm{P}}(j(\mathfrak{a})) = j(\mathfrak{p}^{-1}\mathfrak{a}) \qquad (3.1)$$

since $\mathfrak{P} \not| D$. Since there exists a split prime in every ideal class, this proves that the elements $j(\mathfrak{a}_i)$ are all conjugate over k. Furthermore, (3.1) shows that $\mathfrak{p}$ splits in K if and only if it is principal, since $\mathfrak{p}$ splits completely in

K/k if and only if the Frobenius $\sigma_{\mathfrak{P}}$ for any prime $\mathfrak{P}$ over $\mathfrak{p}$ is trivial, and by (3.1) this happens if and only if $\mathfrak{p}$ is principal, since $\mathfrak{p}$ is principal if and only if $\mathfrak{p}^{-1}$ is principal and the j-function takes the same value on $\mathfrak{p}^{-1}\mathfrak{a}$ and $\mathfrak{a}$ iff these two lattices are multiples of each other, that is if and only if $\mathfrak{p}^{-1}$ is principal. Since this construction can be applied to any prime $\mathfrak{p}$ of k outside a set of density zero, K must be the Hilbert class field of k, and hence $\{j(a_i)\}$ must be a full set of conjugates, since there are as many elements in this set as the class number of k, and this is equal to the degree of the extension K/k.

3.2.2 Ray Class Fields

Now we prove that if F_N is the field of modular functions over $\mathbb{Q}$ of level N, and k is an imaginary quadratic field, then $kF_N(z)$, the field generated over k by values $f(z)$ for $f \in F_N$, $z \in k$ is the ray class field of k to the modulus N. In the following, k will denote an imaginary quadratic field with integer ring $\mathcal{O}$, $\mathfrak{a}$ will denote an ideal of $\mathcal{O}$, and K will denote the Hilbert class field of k.

Lemma 1 *Let A be an elliptic curve having j-invariant $j_A = j(\mathfrak{a})$ and given by a Weierstrass equation $y^2 = x^3 + ax + b$ with $a, b \in k(j_A) \subseteq K$, and discriminant Δ. Let $\phi : \mathbb{C}/\mathfrak{a} \to A_{\mathbb{C}}$ be an analytic representation of A, and define*

$$S = \{2, 3\} \bigcup \{p \in \mathbb{Z} \text{ prime: } p|a \text{ or } p|b \text{ or } p|\Delta\}.$$

Let $\mathfrak{p}$ be a split prime of k whose norm $N(\mathfrak{p}) = p \notin S$, and let $\sigma = \sigma_{\mathfrak{p}}$ be the associated Frobenius element. Then there exists an analytic representation $\psi : \mathbb{C}/\mathfrak{p}^{-1}\mathfrak{a} \to A^{\sigma}_{\mathbb{C}}$ and an isogeny $\lambda : A_{\mathbb{C}} \to A^{\sigma}_{\mathbb{C}}$ (where A^{σ} is defined by $y^2 = x^3 + a^{\sigma}x + b^{\sigma}$) such that

$$\begin{tikzcd}
\mathbb{C}/\mathfrak{a} \arrow[r] \arrow[d] & A_{\mathbb{C}} \arrow[d, "\lambda"] \\
\mathbb{C}/\mathfrak{p}^{-1}\mathfrak{a} \arrow[r] & A^{\sigma}_{\mathbb{C}}
\end{tikzcd}$$

commutes, and such that $\tilde{\lambda}$, the reduction of λ to the curve $\tilde{A}$ defined over $\tilde{K} = K/\mathfrak{P}$ for some fixed P an ideal of $\mathcal{O}_K$ over p is equal to the Frobenius homomorphism $\pi : \tilde{A} \to \tilde{A}^{(p)}$.

Proof: We will denote reduction modulo $\mathfrak{P}$ by a tilde. The proof of theorem 14 gives the diagram shown in the statement of the lemma, and the fact that the reduction of A_K^σ modulo $\mathfrak{p}$, $\tilde{A}^\sigma$, is isomorphic to $\tilde{A}^{(p)}$. First assume that $a \neq 0 \neq b$, then $\mathfrak{p}$ does not divide a or b, and since $\sigma(z) \equiv a\ (mod\,\mathfrak{P})$, the curve $\tilde{A}^\sigma$ is defined by a Weierstrass equation with non-zero linear and constant term, and consequently has as automorphisms only $(x,y) \mapsto (x, \pm y)$ (see section 2.2.3). On the other hand, A^σ also has these two maps as automorphisms, and the automorphisms of A^σ clearly reduce to give the corresponding automorphisms of $\tilde{A}^\sigma$. Knowing this, the lemma follows, since there exists some automorphism $\tilde{\epsilon}$ of $\tilde{A}^\sigma$ such that $\tilde{\lambda} = \tilde{\epsilon} \circ \pi$. But $\tilde{\epsilon} = \pm 1$ is the reduction of the automorphism $\epsilon = \pm 1$ of A^σ, and so we can simply replace λ and ψ in the diagram by $\lambda \circ \epsilon^{-1}$ and $\epsilon^{-1} \circ \psi$ respectively to obtain a diagram satisfying the conditions of the lemma. If a or b is zero, the same argument still applies, since in this case the automorphism groups of A^σ and $\tilde{A}^\sigma$ will either be both order four cyclic or both order six cyclic, and since the residue fields are characteristic not two or three there will be no collapsing of automorphisms in this case either.

Theorem 15 *With notations as above, Let A be an elliptic curve with j-invariant $j_A = j(\mathfrak{a})$ defined over $k(j_A)$ whose ring of endomorphisms is isomorphic to $\mathcal{O}$ (so the lattice $\mathfrak{a}$ is an ideal of $\mathcal{O}$). Let h be the Weber function on $\mathbb{C}/\mathfrak{a}$, as defined in section 2.2, and let A_N be the N-torsion subgroup of A. Then $k(j_A, h(A_N))$ is the ray class field of k to the conductor N.*

Proof: Let K be the smallest Galois extension of k containing j_A and all elements of $h(A_N)$. Take $\mathfrak{p}$ in k a split prime whose norm p is not in the set S defined in lemma 1, and is prime to N. Let $\mathfrak{P}$ be a prime of K over $\mathfrak{p}$. Again the tilde denotes reduction modulo $\mathfrak{P}$. Let σ, ϕ, and ψ be as in lemma 1. The points of finite order of $\mathbb{C}/\mathfrak{a}$ are exactly the points of $k/\mathfrak{a}$; denote the points of order N of this group by $(k/\mathfrak{a})_N$. the map λ on the reduced curve $\tilde{A}$ is equal to the Frobenius map π, and since $(N, p) = 1$ reducing modulo $\mathfrak{P}$ induces an injection $A_N \hookrightarrow \tilde{A}_N$, so that when we restrict all maps to the N-torsion subgroups, $\sigma = \lambda$, and the diagram of lemma 1 becomes

$$\begin{tikzcd}
(k/\mathfrak{a})_N \arrow[r] \arrow[d] & A_N \arrow[d, "\sigma"] \\
(k/\mathfrak{p}^{-1}\mathfrak{a})_N \arrow[r] & A_N^{\sigma}
\end{tikzcd}$$

Since λ has degree p which is prime to N, $\sigma = \lambda$ maps the N-torsion group of A onto the N-torsion group of A^σ, and the group $(k/\mathfrak{a})_N$ onto $(k/\mathfrak{p}^{-1}\mathfrak{a})_N$ (this map is given, rather unenlighteningly, by $[z] \mapsto [z]$). To prove that K is the ray class field to conductor N of k, it suffices to prove that $\mathfrak{p}$ splits completely in K if and only if $\mathfrak{p} = (\alpha)$ for some $\alpha \in k$, $\alpha \equiv 1$ $(mod\ N)$. If $\mathfrak{p}$ satisfies this condition, then p splits completely in the Hilbert class field $k(j_A)$, so $\sigma\mathfrak{p}|_{k(j_A)}$ is the identity (since the restriction to $k(j_A)$ of the Frobenius of a prime ideal of K over $\mathfrak{p}$ is the Frobenius of a prime ideal of $k(j_A)$ over $\mathfrak{p}$), so $A^\sigma = A$ (A is defined over $k(j_A)$). Multiplication by α gives an isomorphism $(k/\mathfrak{p}^{-1}\mathfrak{a}) \overset{\alpha}{\to} (k/\mathfrak{a})$, and composing this map with the diagram above gives the following commutative diagram (where ϕ' is some analytic representation of $A_{\mathbb{C}}$).

$$\begin{tikzcd}
(k/\mathfrak{a})_N \arrow[r] \arrow[d, "\phi"'] & (k/\mathfrak{p}^{-1}\mathfrak{a})_N \arrow[r, "\alpha"] \arrow[d, "\psi"'] & (k/\mathfrak{a})_N \arrow[d, "\phi'"'] \\
A_N \arrow[r, "\sigma"] & A_N \arrow[r, "Id"] & A_N
\end{tikzcd}$$

If $\alpha \equiv 1$ $(mod\ N)$ then the composite map on the top is the identity, since elements of $(k/\mathfrak{a})_N$ can be represented by elements of k with the form c/N with $c \in \mathfrak{a}$, and $\alpha \equiv 1$ $(mod N)$ means that $(\alpha - 1)(c/N) \in \mathfrak{a}$ so the map $\alpha - 1$ is constantly zero on $(k/\mathfrak{a})_N$. ϕ' differs from ϕ by some automorphism of A, so if $\alpha \equiv 1$ $(mod N)$ by the commutativity of the diagram any $P = \phi(t) \in A_N$ differs from $\sigma(P)$ by an automorphism of A, so the Weber functions on the points P and $\sigma(P)$ are equal, since the Weber function gives the same value for two points iff they are in the same orbit under the automorphism group of the curve. But this means that any $P \in A_N$ differs from $\sigma(P)$ by an automorphism of A, and therefore $h(P) = h(\sigma P) = h^{\sigma}(\sigma P) = \sigma(h(P))$, so σ acts trivially on all N-torsion points, and since σ also acts as the identity on $k(j_A)$, σ is the identity automorphism of K, that is $\mathfrak{p}$ splits completely in K.

Conversely, if $\mathfrak{p}$ splits completely in K, then $\mathfrak{p}$ splits completely in the Hilbert class field of k (since this is contained in K), and thus $\mathfrak{p} = (\alpha)$ for some $\alpha \in \mathcal{O}$. Thus we need to show that $\alpha \equiv 1$ $(mod\ N\mathcal{O})$. Now $(k/\mathfrak{a})_N$ is a principal $\mathcal{O}$-module (This is true since $(k/\mathfrak{a})_N$ is an $\mathcal{O}$-module with annihilator $N\mathcal{O}$, and thus is a faithful $\mathcal{O}/N\mathcal{O}$-module - but $|\mathcal{O}/N\mathcal{O}| = |(k/\mathfrak{a})_N| = N^2$, so any element of $(k/\mathfrak{a})_N$ generates the group as a $\mathcal{O}/N\mathcal{O}$-module and hence as a $\mathcal{O}$-module). Let u be a $\mathcal{O}$-module generator of $(k/\mathfrak{a})_N$. Then we have

$$\begin{aligned} h(\phi(u)) &= \sigma(h(\phi(u))) && \text{(since p splits completely, } \sigma = Id \text{ on } k(j_A))\\ &= h^{\sigma}(\phi(u)^{\sigma}) \\ &= h(\phi(u)^{\sigma}) && (h \text{ is defined over } k(j_A))\\ &= h(\psi'(u)) \\ &= h(\phi(\alpha u)) && \text{(by the commutative diagram)} \end{aligned}$$

So $\phi(u)$ and $\phi(\alpha u)$ differ by an automorphism of A. The automorphism group of A is cyclic of order two, four or six, so we must have $\alpha u = \zeta u$ as elements of $k/\mathfrak{a}$ for some second, fourth or sixth root of unity $\zeta \in k$ - so since u generates $(k/\mathfrak{a})_N$ we have $\alpha v \equiv \zeta v$ $(mod\ \mathfrak{a})$ for all $v \in (k/\mathfrak{a})_N$, or $(\zeta^{-1}\alpha - 1) \in N\mathcal{O}$. Since $(\zeta^{-1}\alpha) = (\alpha) = \mathfrak{p}$, this shows that $\mathfrak{p}$ is principal and generated by an element of k congruent to one modulo $N\mathcal{O}$ as required.

Corollary *If F_N is the field of modular functions of weight zero and level N over $\mathbb{Q}$, and $\mathfrak{a}$ is an integral ideal of $\mathcal{O}$, given as a lattice by $[z_1, z_2]$, and $z = z_1/z_2 \in \mathcal{H}$, then the field $kF_N(z)$ generated over k by numbers $f(z)$ for $f \in F_N$ defined at z is the ray class field of k to the modulus N.*

Proof: Denote the ray class field by K. Since the value of the Weber function on the N-torsion subgroup of $\mathbb{C}/\mathfrak{a}$ are exactly the values of the Fricke functions

$$f_{a,b}(z) = h\left(\frac{az+b}{N}\right)$$

for a and b ranging through complete sets of residue classes modulo N , and since these Fricke functions generate the modular function field of level N, by theorem 15 we have $K \subseteq kF_N(z)$. On the other hand, If we denote by R the integral closure of $\mathbb{Z}[j]$ in F_N and by $\mathcal{M}$ the maximal ideal of R consisting of those functions of R which vanish at z, then by the results of section 2.5 on the ramification of ideals in R, we can conclude that any element of $Gal(F_N/\mathbb{Q}(j))$ which fixes all values $f_{a,b}(z)$ is in the inertia group of the ideal $\mathcal{M}$, and consequently fixes the entire residue field; and since $Gal(F_N/\mathbb{Q}(j))$ maps onto $Gal(F_N(z)/\mathbb{Q}(j(z)))$, we conclude that

$$Gal(kF_N(z)/kj(z)) = Gal(K/kj(z)).$$

Both extensions are Galois, so this shows that $kF_N(z) = K$.

3.3 Quadratic Forms and Quadratic Fields

We now elucidate the connection between quadratic field theory and quadratic form theory which is at the root of Mahler's work linking the behaviour of the Dirichlet L-functions associated to imaginary quadratic fields on the real line close to one with the class numbers of the corresponding fields. This is in fact an extension of the famous work of Dirichlet whose class number formula for imaginary quadratic fields gives a (precise) relationship between the value of the L-function at 1 and the class number of the field. By examining the behaviour of the function on an interval near one, Mahler is able to obtain a good bound for the class number as the field discriminant grows large, and conversely.

In order to properly explain this work we must first outline the basic correspondence between quadratic forms and ideals. The idea is quite simple. Let $k = \mathbb{Q}(\sqrt{-d})$ be an imaginary quadratic field ($-d < 0$ a field discriminant) with ring of integers $\mathcal{O}$. Then given any lattice $[\alpha, \beta]$ of $\mathcal{O}$, $N(\alpha x + \beta y) = (N\alpha)x^2 + (\alpha\bar{\beta} + \beta\bar{\alpha})xy + N(\beta)y^2$ (where $\bar{\cdot} : k \to k$ is complex conjugation) is an integral binary quadratic form with discriminant $-d$

(where $N = N_{k/\mathbb{Q}}$ is the norm map). Since all ideals, being (in particular) finite indexed subgroups of $\mathcal{O}$, are lattices we can use this fact to get a correspondence between the ideal classes of $\mathcal{O}$ and appropriately defined isomorphism classes of integral quadratic forms with discriminant $-d$. Throughout, k will denote a fixed imaginary quadratic field of discriminant $-d$, $\mathcal{O}$ its ring of integers, and we will assume that any lattice $[\alpha, \beta] \subseteq \mathcal{O}$ is in fact an ideal of $\mathcal{O}$.

The first problem is that the modules $[\alpha, \beta]$ and $[\bar{\alpha}, \bar{\beta}]$ give rise to the same form. Since $[\alpha, \beta][\bar{\alpha}, \bar{\beta}]$ is a principal ideal, assuming the isomorphism above existed we would be forced to conclude that any ideal class group had exponent two, but this is false (for example, $\mathbb{Z}(\sqrt{-14})$ has the cyclic group of order four as its ideal class group). Hence if we are to gain a correspondence between ideal classes and quadratic forms we must have a way to distinguish $[\alpha, \beta]$ from $[\bar{\alpha}, \bar{\beta}]$. This is done using the concept of order.

Definition: A base $\{\alpha, \beta\}$ of the lattice $[\alpha, \beta]$ is said to be *ordered* if $(\alpha\bar{\beta} - \bar{\alpha}\beta)/\sqrt{-d} > 0$. (This element is in $\mathbb{Q}$ since it is invariant under complex conjugation, and in fact it is in $\mathbb{Z}$, since $\alpha\bar{\beta} - \bar{\alpha}\beta = N([\alpha, \beta])\sqrt{-d}$.)

Now exactly one of $[\alpha, \beta]$ and $[\bar{\alpha}, \bar{\beta}]$ is ordered, and this gives us a method for eliminating the problem arising above - restrict to ordered bases. The next step is to define an equivalence relation on the integral binary quadratic forms of discriminant $-d$ which corresponds to equivalence of ideals. Since bases of a lattice are related by a matrix of $GL_2(\mathbb{Z})$ with determinant ± 1 and multiplying by a matrix of determinant -1 takes an ordered basis to an nonordered basis, we conclude that ordered bases of a given lattice are related by a matrix of $SL_2(\mathbb{Z})$. This explains the following

Definition: Two integral binary quadratic forms F_1 and F_2 are said to be *equivalent* if for some matrix $\begin{pmatrix} P & Q \\ R & S \end{pmatrix} \in SL_2(\mathbb{Z})$ we have $F_1(x, y) = F_2(Px + Qy, Rx + Sy)$.

Definition: An integral binary quadratic form $ax^2 + bxy + cy^2$ is said to be *primitive* if $\gcd(a, b, c) = 1$, and an ideal $[\alpha, \beta]$ is said to be *primitive* if no integer divides α and β in $\mathcal{O}$.

Proposition 1 *(a) If $[\alpha, \beta]$ is an ordered module basis of the $\mathcal{O}$ ideal $\mathfrak{a}$, the quadratic form $N(\alpha x + \beta y)/N(\mathfrak{a})$ is a primitive integral quadratic form of discriminant $-d$ which we will denote by $\phi([\alpha, \beta])$.*

(b) Conversely, given a primitive form $ax^2 + bxy + cy^2$ of discriminant $-d$, $[a, \frac{b-\sqrt{-d}}{2}]$ is an ordered basis of a primitive integral $\mathcal{O}$-ideal $\mathfrak{a}$ which we will denote by $\psi(ax^2 + bxy + cy^2)$.

Proof: The coefficients of the form $F(x, y) = N(\alpha x + \beta y)$ are $N(\alpha)$, $\alpha\bar{\beta} + \beta\bar{\alpha}$ and $N(\beta)$, all of which are integers divisible by $N\mathfrak{a}$ since they are divisible by this element in $\mathcal{O}$ ($\mathbb{Z}$ is integrally closed). Say $N(\alpha x+\beta y)/N(\mathfrak{a}) = ax^2 + bxy + cy^2$. Since this form has discriminant

$$b^2 - 4ac = \frac{(\alpha\bar{\beta} + \bar{\alpha}\beta)^2 - 4N(\alpha\beta)}{(N\mathfrak{a})^2} = \left(\frac{\alpha\bar{\beta} - \bar{\alpha}\beta}{N\mathfrak{a}}\right)^2 = -d,$$

which is a field discriminant, $\gcd(a, b, c) = 1$ if $-d \equiv -1 \ (mod\ 4)$, and $\gcd(a, b, c) \leq 2$ if $-d/4 \equiv 2$ or $3 \ (mod\ 4)$. Assume the latter, and that $\gcd(a, b, c) = 2$. Then $-d/4 = \frac{b^2-4ac}{4} \equiv \left(\frac{b}{2}\right)^2 \equiv 0$ or $1 \ (mod\ 4)$, a contradiction. Thus in all cases $\gcd(a, b, c) = 1$, so $N(\alpha x + \beta y)/N(\mathfrak{a})$ is primitive integral.

(b) Let $\mathfrak{j} = [a, \frac{b-\sqrt{-d}}{2}]$. $\frac{b-\sqrt{-d}}{2} \in \mathcal{O}_k$ if $-d \equiv 1 \ (mod\ 4)$, since in this case b must be odd. On the other hand, if $-d \equiv 0 \ (mod\ 4)$, b is even, so $\frac{b-\sqrt{-d}}{2} \in \mathcal{O}_k$. Hence in either case, the ideal $\mathfrak{j}$ is integral. If an integer n divides both a and $\frac{b-\sqrt{-d}}{2}$, then n must divide $\frac{b+\sqrt{-d}}{2}$, and hence $n|b$, and we also have $n^2|(-d)$, so $n = 2$ since $-d$ is a field discriminant. But now we have that $16|(b^2 - 2b\sqrt{-d} - d) = 2b^2 - 2b\sqrt{-d} - 4ac$ and hence $8|(b^2 - b\sqrt{-d} - 2ac)$, which implies that $4 \nmid (b - \sqrt{-d})$ since $2|b$ and $8 \nmid 2ac$ by primitivity. This is a contradiction, so we conclude that $\mathfrak{j}$ is primitive. A simple computation shows that the module $\mathfrak{j}$ is indeed an ideal, and the given basis is ordered since

$$\frac{a\overline{(b-\sqrt{-d}/2)}-\bar{a}(b-\sqrt{-d}/2)}{\sqrt{-d}}=a>0.$$

Theorem 16 *Let $-d$ be a field discriminant. The maps ϕ and ψ given in the above proposition induce a one to one mutually inverse correspondence between I/P and the set of primitive integral quadratic forms of discriminant $-d$ modulo proper equivalence, where I is the group of fractional $\mathcal{O}$-ideals and P is the subgroup of principal ideals.*

Proof: Calculation shows that $\phi\circ\psi(F)=F$. On the other hand, if $\mathfrak{j}=[\alpha,\beta]\lhd\mathcal{O}$ is an ideal given by an ordered module basis, then

$$\psi\circ\phi(\mathfrak{j})=\left[\frac{N(\alpha)}{N(\mathfrak{j})},\frac{(\alpha\bar{\beta}+\beta\bar{\alpha})}{2N(\mathfrak{j})}-\frac{\sqrt{-d}}{2}\right],$$

and therefore

$$N(\mathfrak{j})(\psi\circ\phi(\mathfrak{j}))=\left[N(\alpha),\frac{(\alpha\bar{\beta}+\beta\bar{\alpha})}{2}-\left(\frac{\alpha\bar{\beta}-\beta\bar{\alpha}}{2}\right)\right]=\bar{\alpha}[\alpha,\beta],$$

so $\psi\circ\phi(\mathfrak{j})$ and $\mathfrak{j}$ are equivalent. This shows that $\psi\circ\phi=Id$ on I/P.

Now it must be proven that the maps respect the equivalence relations so that they induce well defined as maps on the quotients. We'll denote both of these equivalence relations by $\sim$. Suppose that $F_1(x,y)$ and $F_2(X,Y)$ are equivalent under $X\mapsto px+qy$, $Y\mapsto rx+sy$ with $ps-qr=1$, $p,q,r,s\in\mathbb{Z}$. If for $i=1,2$ we have $\mathfrak{j}_i=\psi(F_i)=[\alpha_i,\beta_i]$, then since $\phi\circ\psi=Id$,

$$F_i(x,y)=\frac{(\alpha_i x+\beta_i y)(\bar{\alpha}_i x+\bar{\beta}_i y)}{N(\mathfrak{j}_i)}$$

for $i=1,2$. Now there exist some $A,B\in\mathbb{C}$ such that $\alpha_2 A+\beta_2 B=0$; so by the equivalence of F_1 and F_2 we must have

(3.1) $$\frac{-\beta_1}{\alpha_1}=\frac{a}{b}\quad\text{or}\quad\frac{-\bar{\beta}_1}{\bar{\alpha}_1}=\frac{a}{b},$$

where $\begin{pmatrix} A & B \end{pmatrix} = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} a & b \end{pmatrix}$. We now show that the first of these options implies that $\mathfrak{j}_1 \sim \mathfrak{j}_2$ while the second leads to a contradiction. If the first equality holds,

$$\frac{-\beta_2}{\alpha_2} = \frac{A}{B} = \frac{-\beta_1 p + \alpha_1 q}{-\beta_1 r + \alpha_1 s},$$

so that

$$\frac{\alpha_2}{-\beta_1 r + \alpha_1 s} = \frac{\beta_2}{-\alpha_1 q + \beta_1 p} = \frac{\lambda}{\mu}$$

for some λ and μ, so $\mu\mathfrak{j}_2 = [\mu\alpha_2, \mu\beta_2] = \lambda[-\beta_1 r + \alpha_1 s, \beta_1 p - \alpha_1 q] = \lambda\mathfrak{j}_1$, so since $\det\begin{pmatrix} s & -r \\ -q & p \end{pmatrix} = 1$, $\mathfrak{j}_1 \sim \mathfrak{j}_2$.

On the other hand, if the second equality of (3.1) above holds, following the same steps yields $\mu\mathfrak{j}_2 = \lambda[-\bar{\beta}_1 r + \bar{\alpha}_1 s, \bar{\beta}_1 p - \bar{\alpha}_1 q] = \lambda[\bar{\alpha}_1, \bar{\beta}_2]$. Since $[\alpha_1, \beta_1]$ is ordered, $[\bar{\alpha}_1, \bar{\beta}_1]$ is not, but since $ps - qr = 1$, $[\lambda\bar{\alpha}_1, \lambda\bar{\beta}_1]$ is ordered, which implies that $N(\lambda) < 0$, which is a contradiction since all elements in imaginary quadratic fields have positive norm. Thus $F_1 \sim F_2$ implies $\psi F_1 \sim \psi F_2$.

Now assume that for $i = 1, 2$, $\mathfrak{j}_i = [\alpha_i, \beta_i]$ are ordered bases, and that $\lambda\mathfrak{j}_1 = \mu\mathfrak{j}_2$ for some elements λ and μ. Then $\lambda\alpha_1 = p\mu\alpha_2 + q\mu\beta_2$ and $\lambda\beta_1 = r\mu\alpha_2 + s\mu\beta_2$ for some matrix $\begin{pmatrix} p & q \\ r & s \end{pmatrix} \in SL_2(\mathbb{Z})$, so $N(\lambda\alpha_1 x + \lambda\alpha_2 y) \sim N(\mu\alpha_2 x + \mu\beta_2 y)$ (as quadratic forms). But the construction of the form $\phi(\mathfrak{j})$ involved dividing by the norm of $\mathfrak{j}$, so $\mathfrak{j}$ and $c\mathfrak{j}$ give rise to the same image under ϕ. Thus $\phi(\mathfrak{j}_1) = N(\alpha_1 x + \beta_1 y) \sim N(\alpha_2 x + \beta_2 y) = \phi(\mathfrak{j}_2)$, so that $\mathfrak{j}_1 \sim \mathfrak{j}_2$ implies $\phi(\mathfrak{j}_1) \sim \phi(\mathfrak{j}_2)$. Hence these two maps prove a bijection between the specified sets.

Remark: It is of course possible to use the maps ϕ and ψ above to 'push' the multiplication on I/P to the set of quadratic forms of discriminant $-d$ modulo proper equivalence, and hence give this set the structure of an abelian group. It turns out that there is a 'natural' multiplication on this set which is the same as this transported multiplication, so that the maps ϕ and ψ in fact are group isomorphisms. The multiplication on the quotient set of quadratic forms is outlined in [3].

We have only gone through the construction of the maps ϕ and ψ in the case where $-d < 0$. In the case of real quadratic fields, the situation is more complicated, since there can be elements with negative norm, so that the definition of equivalence of ideals must be tightened in order to match up exactly with the notion of proper equivalence of quadratic forms.

The correspondence outlined above can also be expressed analytically in terms of equality of certain zeta functions. If $-d < 0$ is the discriminant of the imaginary quadratic field K with integer ring $\mathcal{O}$, and $\mathcal{C} \in I/P$ an ideal class of $\mathcal{O}_{\mathbb{Q}(\sqrt{-d})}$, the zeta function of $\mathcal{C}$ is defined as

$$\zeta_{\mathcal{C}}(z) = \sum_{\mathfrak{a}\in\mathcal{C}} \frac{1}{(N\mathfrak{a})^z},$$

the sum over all integral ideals a in the ideal class. Then we clearly have

$$\zeta_K = \sum_{\mathcal{C}\in I/P} \zeta_{\mathcal{C}}.$$

Now if we define the zeta function of a (primitive integral) quadratic form $Q(x,y) = ax^2 + bxy + cy^2$ to be

$$\zeta_Q(s) = \frac{1}{U(-d)} \sum_{m,n\in Z}{}' \frac{1}{(am^2 + bmn + cn^2)^s},$$

where $U(-d)$ is the number of units in K (so that $U(-d) = 2$ if $d > 4$, $U(-3) = 6$ and $U(-4) = 4$) then we have the following

Proposition 2 *If the ideal class $\mathcal{C}$ corresponds under the above bijection to an equivalence class of quadratic forms containing the form Q, then $\zeta_{\mathcal{C}} = \zeta_Q$.*

Proof: Let $[\alpha, \beta] = \mathfrak{a}$ be an ordered module basis for $\mathfrak{a} \in \mathcal{C}$. Then since properly equivalent quadratic forms take exactly the same set of values, we can take Q to be $\phi(\mathfrak{a})$. Now there exists m, n such that $Q(m,n) = N(\mathfrak{a})$. Indeed, since

$$\phi(\mathfrak{a}) = \frac{N(\alpha x + \beta y)}{N(\mathfrak{a})},$$

it suffices to find an element $a \in \mathfrak{a}$ such that $N(a) = N(\mathfrak{a})^2$, and to show this is possible for any $\mathfrak{a}$, it suffices to show it for prime ideals $\mathfrak{p}$. Say $\mathfrak{p}$ is above the prime p of $\mathbb{Z}$. If $\mathfrak{p}$ is split, then we can take p as our required element, and if $\mathfrak{p}$ is inert or ramified we can take the element p^2. This shows that Q takes all integer values corresponding to norms of ideals of $\mathcal{C}$. On the other hand, factoring the form Q in the field K shows that if $d > 4$, Q takes each value zero times or twice, and if $d = -3$ each value zero or six times and if $d = -4$ zero or four times. Thus the two zeta functions are indeed equal.

We end this section with a proposition which is used to show the existence of certain 'canonical' integral quadratic forms within each equivalence class.

Proposition 3 *Given a positive definite form $ax^2+bxy+cy^2$ of discriminant $-d = b^2 - 4ac$, there exists a properly equivalent form with $0 < a \leq \sqrt{\frac{d}{3}}$.*

Proof: It suffices to find an equivalent form $Ax^2 + Bxy + Cy^2$ with $C > A > |B|$, since in this case we have $d = 4AC - B^2 \geq 3A^2$ as required. We can assume without loss of generality that $c < a$, since the matrix $\begin{pmatrix} 1 & 0 \\ N & 1 \end{pmatrix}$ sends $Ax^2 + Bxy + Cy^2$ to $(A + BN + CN^2)x^2 + (B + 2NC)xy + Cy^2$. Now the matix $\begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix} \in SL_2(\mathbb{Z})$ sends $ax^2 + bxy + cy^2$ to $ax^2 + xy(2aN + b) + y^2(aN^2 + bN + c)$. Pick $N > 0$ such that

$$aN^2 + bN + c \geq a \tag{3.2}$$

$$a(N-1)^2 + b(N-1) + c < a, \tag{3.3}$$

then manipulating (2) and (3) gives $2aN+b > 0$, and $aN^2+bN+c < 2aN+b$, so we have a form $Ax^2 + Bxy + Cy^2$ with $0 < C < B$, $C > A$, Pick $X > 0$ such that $|B - 2XC| \leq C$ ($X > 0$ since $C < B$). Since $\begin{pmatrix} 0 & -1 \\ -1 & X \end{pmatrix} \in SL_2(\mathbb{Z})$ takes $Ax^2 + Bxy + Cy^2$ to $Cx^2 + (B - 2XC)xy + (A + XB + X^2C)y^2$, and $A + XB + X^2C > C$ $(X > 0, A, B \geq 0)$ we have the desired equivalent form.

In section 2.1.2 we proved that the j-function takes the same value on two (lattices representing) ideals of $\mathcal{O}$ exactly when the ideals are in the same ideal class, and since the ideal corresponding to a quadratic form $ax^2+bxy+cy^2$ with discriminant $-d$ is $\mathfrak{a} = [a, \frac{b-\sqrt{-d}}{2}]$, $j(\mathfrak{a}) = j\left(\frac{b-\sqrt{-d}}{2a}\right)$, so by the above propositions we can take the $h(-d)$ distinct values of j on $\mathcal{O}_{\mathbb{Q}(\sqrt{-d})}$-ideals to be $\left\{j\left(\frac{b_i-\sqrt{-d}}{2a_i}\right)\right\}$ $(i = 1, 2, \cdots, h(-d))$ where each $a_i \leq \left|\sqrt{\frac{-d}{3}}\right|$.

3.4 L-functions and Class numbers

We now show the link between the zeroes of the function $L_d = L(s, \chi)$ (where χ is the real primitive character to the modulus d) and the class number of the field $\mathbb{Q}(\sqrt{-d})$ for $-d < 0$ a field discriminant. The technique basically involves bounding the size of the zeta functions of imaginary quadratic number fields uniformly for varying s and d, and for this we use the zeta function ζ_Q of a binary integral quadratic form Q introduced in the last section. In this section we are following [23] and [18]. We begin with a lemma about the Γ-function.

Lemma 1 *Put* $B(u,v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$. *Then*

$$B(u,v) = \int_0^\infty \frac{t^{u-1}dt}{(1+t)^{u+v}} = 2\int_0^\infty \frac{x^{2u-1}}{(1+x^2)^{u+v}}dx.$$

Proof: The substitution $t = x^2$ shows that the two integrals above are equal, and the substitution $t = \frac{y}{1-y}$ shows that the left hand integral above is equal to

$$\int_0^1 t^{u-1}(1-t)^{v-1}dt.$$

On the other hand,

$$\begin{aligned}\Gamma(u)\Gamma(v) &= \int_0^\infty\int_0^\infty e^{-(x+y)}x^{u-1}y^{v-1}dxdy \\ &= 2\int_0^\infty\int_0^\infty e^{-r}r^{u+v-1}(\cos\theta)^{2u-1}(\sin\theta)^{2v-1}drd\theta \\ &= 2\Gamma(u+v)\int_0^{\frac{\pi}{2}}(\cos\theta)^{2u-1}(\sin\theta)^{2v-1}d\theta \\ &= \Gamma(u+v)\int_0^1(1-t)^{u-1}t^{v-1}dt\end{aligned}$$

on making the substitutions $x = r\cos^2\theta$, $y = r\sin^2\theta$, followed by $t = \sin^2\theta$. This proves the lemma.

Now let K be an imaginary quadratic field with discriminant $-d$. We study the zeta function of K by studying the zeta functions of integral quadratic forms of discriminant $-d$. Let $Q(x,y) = ax^2 + bxy + cy^2$ be such a form. From the last section,

$$(3.1) \qquad U(-d)\zeta_Q(s) = \sum_{m,n\in Z}{}' \frac{1}{(am^2+bmn+cn^2)^s}.$$

The terms in the sum with $n = 0$ give $2a^{-s}\zeta(2s)$ (where ζ is the usual Riemann zeta function). Applying the Poisson summation formula (see [14])

$$(3.2) \qquad \sum_{m=-\infty}^{\infty} f(m) = \sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)e^{2\pi imx}dx$$

(with $f(m) = am^2 + bmn + cn^2$ for each value of n) to the terms with $n \neq 0$ gives

$$(3.3) \quad U(-d)\zeta_Q(s) = 2a^{-s}\zeta(2s) + 2\sum_{n=1}^{\infty}\sum_{m=-\infty}^{\infty}{}' \int_{-\infty}^{\infty}\frac{e^{2\pi imx}dx}{(ax^2+bxn+cn^2)^s}$$

The $m = 0$ term in (3.3) is

$$\begin{aligned} &= 2\sum_{n=1}^{\infty}\int_{-\infty}^{\infty}\frac{dx}{(ax^2+bxn+cn^2)^s} \\ &= a^{2s-1}d^{\frac{1}{2}-s}2^{-2s}\zeta(2s-1)\int_{-\infty}^{\infty}\frac{dx}{(x^2+1)^s},\end{aligned}$$

on making the substitution $x = nu$ followed by completing the square in the denominator and making the substitution $u = \frac{\sqrt{d}}{2a}(x - \frac{b}{2a})$. Now we apply the lemma with $u = \frac{1}{2}$, $v = s$ and have that the $m = 0$ part of (3.3) is

$$\frac{a^{2s-1}d^{\frac{1}{2}-s}2^{-2s}\Gamma(s-\frac{1}{2})\sqrt{\pi}}{\Gamma(s)}.$$

Now the $m \neq 0$ part of (3.3) is

$$2\sum_{n=1}^{\infty}\sum_{m\in Z\setminus\{0\}}\int_{-\infty}^{\infty}\frac{e^{2\pi ixm}dx}{(ax^2+bxn+cn^2)^s}.$$

Call the term corresponding to (m, n) in this sum, T_{mn}. Then a computation similar to the above shows that

$$T_{mn} = \frac{a^{s-1}4^{s-\frac{1}{2}}}{n^{2s-1}d^{s-\frac{1}{2}}}e^{-2\pi ib/a}I(\alpha),$$

where

$$I(\alpha) = \int_{-\infty}^{\infty}\frac{e^{2\pi i\alpha z}dz}{(z^2+1)^s}, \qquad \alpha = \frac{mn\sqrt{d}}{2a}.$$

Thus in order to understand the behaviour of $\zeta_Q(s)$ we must understand the behaviour of the integral $I(\alpha)$. Since we proved in the previous section that every binary integral quadratic form was $SL_2(\mathbb{Z})$ equivalent to a reduced form with $\sqrt{d} \geq a\sqrt{3}$, we can assume that our form Q has this property, and hence that $\alpha \geq \frac{\sqrt{d}}{2}$. In this case we have the following

Lemma 2 *If* $\alpha \geq \frac{\sqrt{d}}{2}$, $I(\alpha) = O(e^{-\pi\alpha})$ *uniformly in* s, *for* s *in compact subsets of* $\{z : \Re(z) \geq \frac{1}{2}\}$.

Proof: Let $\sigma = \Re(s)$. We will use a branch of $(z^2+1)^s$ analytic in $\mathbb{C} \setminus (\{z : \Re z = 0, \Im z \leq -1\} \bigcup \{z : \Re z \geq 0, \Im z = \frac{1}{2}\})$. Let $\frac{e^{2\pi i\alpha z}}{(z^2+1)^s} = F$. Then $I(\alpha)$ is equal to any sum of the form $I_1 + I_{2,\delta} + I_{3,\delta} + I_{4,\delta}$, where

$$I_1 = \lim_{r\to\infty}\int_{\gamma_{1,r}}Fdz, \qquad I_{2,\delta} = \int_{\gamma_{2,\delta}}Fdz,$$

$$I_{3,\delta} = \int_{\gamma_{3,\delta}}Fdz, \qquad I_{4,\delta} = \int_{\gamma_{4,\delta}}Fdz,$$

with $\gamma_{1,r}$ the upper half of a circle of radius r centred at the origin, traversed clockwise, γ_2 the circle of radius $\frac{1}{2}$ centred at i and traversed counterclockwise, $\gamma_{2,\delta}$ is $\gamma_2 \setminus \{z : 1-\delta \leq \Im(z) \leq 1+\delta\}$, $\gamma_{3,\delta}$ is that part of the line of constant imaginary part $1-\delta$ to the right of the circle γ_2, traversed from left to right, and $\gamma_{4,\delta}$ is that part of the line of constant imaginary part $1+\delta$ to the right of the circle γ_2 and traversed from right to left. This is since by choice of branch, F is analytic in the interior of the region traced by these four curves and the real axis (which is the path of integration of $I(\alpha)$). We evaluate each of these integrals separately.

$I_1 = 0$ since putting $z = re^{i\theta}$ into Fdz shows that the denominator of the integrand is approximately r^σ for large r, while the integral of the absolute value of the numerator is

$$\int_0^\pi e^{-2\pi\alpha \sin\theta} r d\theta \leq 2 \int_0^{\frac{\pi}{2}} e^{-4\alpha r\theta} r d\theta,$$

since $\pi \sin\theta \geq 2\theta$ on $[0, \frac{\pi}{2}]$. The integral on the right hand side is $\leq 1/2\alpha \leq 1/\sqrt{3}$.

$I_2 = O(e^{-\pi\alpha})$ since putting $z = i + \frac{e^{i\theta}}{2}$ into Fdz and taking absolute value gives

$$\frac{e^{-2\pi\alpha - \pi\alpha\sin\theta}}{2^{-\sigma} e^{i\theta\sigma} |z+i|^\sigma} = O(e^{-\pi\alpha}),$$

and the path of integration has finite length.

To show $I_{3,\delta}$ is $O(e^{-\pi\alpha})$ for $\delta > 0$ sufficiently small, we first change the branch of $(z^2+1)^s$ to one which is analytic on the line $\gamma_3 = \{z : \Im z = 1, \Re z \geq 1\}$ and also on $\gamma_{3,\delta}$ δ sufficiently small (for example, there are branches of this function analytic on $\mathbb{C} \setminus [-i, i]$). This is justified since any two branches of this function differ by a constant of absolute value one on their common domain of definition. For this new branch of F, the absolute value of the integral on γ_3 can be made arbitrarily close to that of the integral on $\gamma_{3,\delta}$ by taking δ sufficiently small (since the integrand is now an analytic function on regions bounded by below and above by γ_3 and $\gamma_{3,\delta}$, on the left by γ_2 and on the right by lines of the form $\{z : \Re z = M\}$, and the integrand can be made arbitrarily small on either of these boundaries by taking δ small enough. Now it is enough to show that the integral

$$\int_{\gamma_3} F dz$$

is $O(e^{-\pi\alpha})$. Put $z = i + \mu$, and the integral in question is

$$e^{-2\pi\alpha} \int_{\frac{1}{2}}^{\infty} \frac{e^{2\pi i\alpha\mu} d\mu}{(2i\mu + \mu^2)^s} = \frac{e^{\pi i\alpha}}{(2\pi i\alpha(i + \frac{1}{4})^s)} - \frac{1}{2\pi i\alpha(-s)} \int_{\frac{1}{2}}^{\infty} \frac{e^{2\pi i\alpha\mu}(2i + 2\mu)}{(2i\mu + \mu^2)^s}$$

(integration by parts). The integral on the right hand side is

$$O\left(\int_{\frac{1}{2}}^{\infty} \frac{d\mu}{\mu^{2\sigma+1}}\right)$$

which is $O(1)$ since $\sigma \geq \frac{1}{2}$. This completes the proof of the lemma.

We now apply this lemma to the bounding of the $m \neq 0$ part of (3.3).

$$\begin{aligned} \left|\sum_{n=1}^{\infty} \sum_{m \in Z}{}' \; T_{mn}\right| &\leq \sum_{n=1}^{\infty} \sum_{m \in Z}{}' \; |T_{mn}| \\ &= O\left(\sum_{n=1}^{\infty} \sum_{m \in Z}{}' \; \frac{a^{\sigma-1} e^{-\pi mn\sqrt{d}/2a}}{d^{\sigma-1/2}} n^{2\sigma-1}\right) \\ &= \frac{a^{\sigma-1}}{d^{\sigma-\frac{1}{2}}} O\left(\sum_{n=1}^{\infty} \frac{1}{n^{2\sigma-1}} \sum_{m=1}^{\infty} (e^{-\pi n\sqrt{d}/2a})^m\right) \\ &= \frac{a^{\sigma-1}}{d^{\sigma-\frac{1}{2}}} O\left(\sum_{n=1}^{\infty} \frac{1}{n^{2\sigma-1}} \frac{e^{-\pi n\sqrt{d}/2a}}{1 - e^{-\pi n\sqrt{d}/2a}}\right), \end{aligned}$$

and the sum is $O(1)$ for $\sigma \geq \frac{1}{2}$, independent of d. Combining this with the $m = 0$ and $n = 0$ pieces of (3.3) gives

(3.4)

$$U(-d)\zeta_Q(s) = 2a^{-s}\zeta(2s) + 2d^{\frac{1}{2}-s}a^{s-1}\left(\frac{\zeta(2s-1)\Gamma(s-\frac{1}{2})\sqrt{\pi}}{\Gamma(s)} + O(1)\right)$$

with the constant $O(1)$ independent of d, uniformly for s in compact subsets of $\{z : \Re(z) \geq \frac{1}{2}\}$. Now we fix $-d < 0$ a field discriminant, and add (3.4) over a full set of reduced forms of discriminant $-d$ to obtain

(3.5)

$$U(-d)\zeta(s)L_d(s) = f_d(s)\zeta(2s) + d^{\frac{1}{2}-s}f_d(1-s)\left(\frac{\zeta(2s-1)\Gamma(s-\frac{1}{2})\sqrt{\pi}}{\Gamma(s)} + g_d(s)\right),$$

where $g_d(s)$ is a positive real function of s, uniformly bounded for all field discriminants d and $s \in [\frac{1}{2}, 1]$, and

$$f_d(s) = \sum a^{-s},$$

the sum being over a full set of reduced integral binary quadratic forms $ax^2 + bxy + cy^2$ of discriminant $-d$.

Lemma 3 *There exists $\sigma_0 \in (\frac{1}{2}, 1)$ and positive constants c_1, c_2, c_3, c_4 such that for all field discriminants $-d < 0$ there exist functions $A_d(s)$ and $B_d(s)$ satisfying $c_1 \le A_d(s) \le c_2$ and $c_3 \le B_d(s) \le c_4$ and*

$$\zeta(s)L_d(s) = A_d(s)f_d(s) - \frac{B_d(s)d^{\frac{1}{2}-s}f_d(1-s)}{1-s} \tag{3.6}$$

for all $s \in (\sigma_0, 1)$.

Proof: Define

$$A_d(s) = \frac{\zeta(s)}{U(-d)} \quad B_d(s) = \frac{(s-1)}{U(-d)}\left(\frac{\zeta(2s-1)\Gamma(s-\frac{1}{2})\sqrt{\pi}}{\Gamma(s)} + g_d(s)\right).$$

Lemma 1 and easy bounding of $\sqrt{t}(1+t)$ on $[0,1]$ and $[1,\infty)$ show that if $s \in (1/2, 1]$,

$$\frac{\Gamma(s-\frac{1}{2})\sqrt{\pi}}{\Gamma(s)} = \int_0^\infty \frac{dt}{\sqrt{t}(1+t)^s} \ge \int_0^\infty \frac{dt}{\sqrt{t}(1+t)} \ge \frac{1}{2}\int_0^1 dt + \frac{1}{2}\int_1^\infty \frac{dt}{t^{3/2}} \ge \frac{3}{2}.$$

Choose $M > 0$ so that for all field discriminants $-d < 0$ and all $s \in [1/2, 1]$, $|g_d(s)| < M$. Since $\zeta(2s-1)$ has a pole at one there exists some $\sigma_0 \in (\frac{1}{2}, 1)$ so that if $s \ge \sigma_0$, $\zeta(2s-1) > 2M/3$ so on $[\sigma_0, 1]$, $A_d(s)$ and $B_d(s)$ are both continuous and nonvanishing. Thus the constants exist, and clearly the functions A_d and B_d defined in this way satisfy (3.6).

Now the class number $h = h(-d)$ is equal to $f_d(0)$. Put

$$\eta = f_d(1) = \sum a^{-1},$$

with the sum as usual over a full set of reduced forms of discriminant $-d$. For $s \in [\sigma_0, 1]$, we have

$$hd^{\frac{1}{2}(1-s)} \leq f_d(1-s) \leq h$$

$$\eta \leq f_d(s) \leq \eta d^{\frac{1}{2}(1-s)}$$

The integral representations of $\Gamma(s)$ and $\zeta(s)$ valid in $0 < \Re(s) < 1$,

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} dt, \qquad \zeta(s) = \int_0^\infty \left(\frac{1}{e^t - 1} - \frac{1}{t} \right) t^{s-1} dt$$

show that $\zeta(s)$ is real and strictly negative for $s \in (0,1)$. The main theorem is now essentially a corollary of the above lemmas.

Theorem 17 (Mahler) *Suppose that a constant $A > 0$ exists such that for all field discriminants $-d < 0$,*

$$h(-d) \geq A\eta \frac{\sqrt{d}}{\log d},$$

where η is as above. Then another constant $B = B(A)$ exists such that no L-function $L_d(s)$ (for $-d$ a field discriminant) has a zero in

$$1 - \frac{B}{\log d} \leq s \leq 1.$$

Proof: Rearranging the above inequalities shows that

$$\frac{(1-s)\zeta(s)L_d(s)}{B_d(s)f_d(s)} = \frac{A_d(s)}{B_d(s)}(1-s) - \frac{f_d(1-s)}{f_d(s)} d^{\frac{1}{2}-s} \leq \frac{c_2}{c_3}(1-s) - \frac{h}{\eta\sqrt{d}}.$$

Therefore on $(\sigma_0, 1)$, since $\zeta(s)$ is negative, we have that

$$L_d(s) \geq \frac{B_d(s)f_d(s)}{(1-s)(-\zeta(s))} \left(\frac{h}{\eta\sqrt{d}} - \frac{c_2}{c_3}(1-s) \right).$$

The bracketed on the right hand side of this inequality is

$$\left(\frac{h}{\eta\sqrt{d}} - \frac{c_2}{c_3}(1-s)\right) \geq \frac{A\sqrt{d}}{\log d} + \frac{c_2}{c_3}(s-1),$$

which will be positive if

$$s \geq \max\left(\sigma_0, 1 - B\frac{\sqrt{d}}{\log d}\right)$$

where $B = c_3 A/c_2$.

3.5 *abc* Implies No 'Siegel Zeroes'

In this section we present the proof that the uniform abc-conjecture for number fields implies the non-existence of 'Siegel zeroes' for L-functions associated to imaginary quadratic number fields. The proof uses the functions $\sqrt{j-1728}$ and $j^{1/3}$ which were introduced in section 2.4 and which we will call γ_3 and γ_2 respectively. By their definition, for every $z \in \mathcal{H}$, these functions give a solution $(x,y) = (\gamma_3(z), \gamma_2(z))$ of the Diophantine equation

$$y^3 = x^2 + 1728. \tag{3.1}$$

Weber proved in [32] that if the class number h of the imaginary quadratic field $k = \mathbb{Q}(\sqrt{d})$ is one then $\gamma_3(\tau)/\sqrt{d}$ and $\gamma_2(\tau)$ are rational integers. Since it was proven in 1.3 that the abc-conjecture implies (3.1) has only finitely many solutions in $\mathbb{Z}$, this shows that the abc-conjecture implies that there exist only finitely many imaginary quadratic fields of class number one. Of course this is of little use since it is already known to be true, but it points in a certain direction. Since γ_3 and γ_2 are both in F_6, the field of modular functions of level six, whenever $z \in \mathcal{H} \bigcap k$ we must have that $\gamma_2(z)$ and $\gamma_3(z)$ are both in L, the field generated over k by all values $f(z)$ for $f \in F_6$ defined at z, which was shown in section 3.2.2 to be the ray class field of k to the modulus six. Furthermore, since $j(z)$ is an algebraic integer for $z \in \mathcal{H} \bigcap k$, $\gamma_2(z)$ and $\gamma_3(z)$ must be algebraic integers as well.

To apply the uniform *abc*-conjecture to (3.1) in the ray class field L we must first bound the size of the normalized discriminant $\Delta_{L/\mathrm{Q}}$ (see section 1.1 for definition).

Lemma 1 $|\Delta_{L/\mathrm{Q}}| \leq 6\sqrt{d}$

Proof: We calculate this discriminant using the Different $D = Diff(L/\mathbb{Q}) \lhd \mathcal{O}_L$ (see [16] for definition and properties). Let N denote the norm from L to $\mathbb{Q}$, and the symbol $||$ mean 'divides exactly' - that is $p^b \mid\mid a$ means $p^b|a$ and $p^{b+1} \not| a$. If $\mathfrak{P} \lhd \mathcal{O}_L$ is a prime of $\mathcal{O}_L$ above p where $p \not| 6d$ then since p is unramified in $L/\mathbb{Q}$, it follows that $\mathfrak{P} \not| D$. If $p|6$ then only a power r of $\mathfrak{P}$ less than or equal to the ramification index $e(\mathfrak{P}/p)$ divides D. Say there are g primes of $\mathcal{O}_L$ above p and that the residue degree is f. Then the p-power part of the discriminant is given by

$$|\prod_{\mathrm{P}|p} N(\mathfrak{P}^r)| = p^{rfg} \leq p^{efg} = p^{[L:\mathbb{Q}]}.$$

If $p|d$, $p \not| 6$ then (putting $\mathfrak{p}$ a prime of $\mathcal{O}_k$ above p and below $\mathfrak{P}$) $\mathfrak{p}$ is unramified in L/k and $p \neq 2$, $\mathfrak{p} \mid\mid Diff(k/\mathbb{Q})$ (since if $\mathfrak{p}^2|Diff(k/\mathbb{Q})$ then $p^2|d$). Since the different is multiplicative this implies that the p part of the different D is $\mathfrak{p}$ (considered as an ideal of $\mathcal{O}_L$) so the p part of the discriminant is

$$N(\mathfrak{p}) = N_{\mathbb{Q}}^k N_k^L(\mathfrak{p}) = N_q^k(\mathfrak{p}^{[L:k]}) = p^{[L:k]} = p^{[L:\mathbb{Q}]/2}.$$

Thus the discriminant $|Disc(L/\mathbb{Q})| \leq (6\sqrt{d})^{[L:\mathbb{Q}]}$, and since the normalized discriminant is the $[L:\mathbb{Q}]^{th}$ root of this quantity, this proves the lemma.

Theorem 18 (Granville/Stark) *The uniform abc-conjecture for number fields implies that there exists some constant $C > 0$ such that for $-d$ a field discrimimant,*

$$h(-d) \geq \frac{\pi\sqrt{d}}{(3A+\epsilon)\log d}\left(\sum \frac{1}{a} + \frac{C}{\log d}\right),$$

where $h(-d)$ is the class number of the imaginary quadratic field $\mathbb{Q}(\sqrt{-d})$ and the sum is over a full set of reduced binary integral quadratic forms $ax^2 + bxy + cy^2$.

Proof: We apply the uniform *abc*-conjecture to solutions $(\gamma_3(\tau), \gamma_2(\tau))$ of (3.1) in L ($\tau \in k$). Applying the conjecture and using lemma 1 to bound the discriminant gives

$$(3.2) \qquad H(\gamma_2(\tau)^3, \gamma_3(\tau)^2, 1728) \ll_\epsilon (\sqrt{d})^A G(\gamma_2(\tau)^3, \gamma_3(\tau)^2, 1728)^{1+\epsilon}.$$

From the definition of the conductor we also have

$$\begin{aligned} G(\gamma_2(\tau)^3, \gamma_3(\tau)^2, 1728) &= G(\gamma_2(\tau), \gamma_3(\tau), 6) \\ &\ll G(\gamma_2(\tau), \gamma_3(\tau), 1) \\ &\leq G(\gamma_2(\tau), 1) N(\gamma_3(\tau), 1), \end{aligned}$$

and for any $\alpha \in \mathcal{O}_L$ the product formula (see [16])

$$\begin{aligned} 1 &= \prod_v \|\alpha\|_v = \prod_{\sigma : L \hookrightarrow \mathrm{C}} \|\alpha\|_\sigma \prod_{\mathrm{p}} \|\alpha\|_{\mathrm{p}}^{v\mathrm{p}(\alpha)} \\ &= N_{\mathrm{Q}}^L(\alpha) \prod_{\mathrm{p}} \|\alpha\|_{\mathrm{p}}^{v\mathrm{p}(\alpha)} \end{aligned}$$

(the product is over maximal ideals $\mathfrak{p} \lhd \mathcal{O}_L$, with $v_{\mathrm{p}}(\alpha)$ the $\mathfrak{p}$-adic valuation) shows that

$$H(\alpha, 1) \geq N_{\mathrm{Q}}^L(\alpha) = \prod_{\mathrm{p}} \|\mathfrak{p}\|_{\mathrm{p}}{}^{-v\mathrm{p}(\alpha)} \geq \prod_{\mathrm{p}} \|\mathfrak{p}\|^{-1} \geq G(\alpha, 1),$$

so since $\gamma_2(\tau)$ and $\gamma_3(\tau)$ are in $\mathcal{O}_L$,

$$\begin{aligned} G(\gamma_2(\tau)^3, \gamma_3(\tau)^2, 1728) &\ll H(\gamma_2(\tau), 1) H(\gamma_3(\tau), 1) = H(\gamma_2(\tau)^3, 1)^{\frac{1}{3}} H(\gamma_3(\tau)^2, 1)^{\frac{1}{2}} \\ &\ll H(\gamma_2(\tau)^3, \gamma_3(\tau)^2, 1728)^{\frac{1}{3}} H(\gamma_2(\tau)^3, \gamma_3(\tau)^2, 1728)^{\frac{1}{2}} \\ &= H(\gamma_2(\tau)^3, \gamma_3(\tau)^2, 1728)^{\frac{5}{6}}. \end{aligned}$$

Putting this into (3.2) and solving for $H(\gamma_2(\tau)^3, \gamma_3(\tau)^2, 1728)$ we have

$$(3.3) \qquad H(\gamma_2(\tau)^3, \gamma_3(\tau)^2, 1728) \ll_\epsilon d^{\frac{3A}{1-5\epsilon}} \ll_\epsilon d^{3A(1+10\epsilon)} = d^{3A+\epsilon}$$

for small ϵ (the last inequality is achieved by replacing ϵ by $\epsilon/30A$). Notice that the *uniform abc*-conjecture is really being used here. Now

$$\gamma_2(\tau)^3 = j(\tau) = \frac{1}{q} + 744 + \sum_{n=1}^{\infty} c_n q^n$$

(where $q = e^{2\pi i \tau}$) so

$$\max\{1, |j(\tau)|\} = \frac{1}{|q|} \max\{|q|, |1 + 744q + \cdots|\} \gg \frac{1}{|q|},$$

since when $|q|$ is very small, $|1 + 744q + \cdots|$ is bounded away from zero because $j(z)$ has a pole at infinity. $|1/q| = e^{\pi\sqrt{d}/a}$ so we have

$$\prod e^{\pi\sqrt{d}/a} \ll \prod \max\{|j(\tau^*)|, 1\} \ll H(\gamma_2(\tau)^3, \gamma_3(\tau)^2, 1728) \ll_\epsilon d^{(3A+\epsilon)h}$$

(with the product over a full set of non-equivalent reduced forms of discriminant $-d$ and τ^* is the element $\frac{-b+\sqrt{-d}}{2a}$ corresponding to the form $ax^2 + bxy + cy^2$) and taking logarithms gives (for some $C = C_\epsilon$)

$$\pi\sqrt{d} \sum \frac{1}{a} \leq (3A + \epsilon) h \log(d) + C,$$

which gives the lower bound on the class number h.

Corollary *The uniform abc-conjecture for number fields implies that there exists some constant $B > 0$ such that for any primitive real Dirichlet character to the modulus $-d$ ($-d$ a field discriminant) the Dirichlet L-function $L(s) = L(s, \chi)$ has no zeroes on the strip*

$$1 - \frac{B}{\log d} \leq s \leq 1.$$

Proof: This is immediate from the above theorem and the result of Mahler from the previous section.

Chapter 4

abc and the Wieferich Criterion

4.1 *abc* and the Wieferich Criterion

In 1909, A. Wieferich [34] proved that if a prime p satisfies

$$(4.1) \qquad 2^{p-1} \not\equiv 1 \ (mod\, p^2)$$

then the first case of Fermat's last theorem holds for p. This result has been generalized by various authors, most recently Granville [9] who proved that the first case of Fermat's last theorem holds for any prime p which satisfies $q^{p-1} \not\equiv 1 \ (mod\, p^2)$ for some prime $q \leq 89$. Despite the fact that the only two primes known to satisfy $2^{p-1} \equiv 1 \ (mod\, p^2)$ are 1093 and 3511, it is not known whether or not there exist infinitely many primes p satisfying (4.1). In 1986, Joseph Silverman [28] proved that the *abc*-conjecture implies that for any $\alpha \in \mathbb{Q}^*$, $\alpha \neq \pm 1$, the number of primes $p \leq X$ satisfying $\alpha^{p-1} \not\equiv 1 \ (mod\, p^2)$ is $\gg \log X$. In this section we show that Silverman's methods can be modified to show that the uniform *abc*-conjecture for number fields implies a similar result for certain elements of real abelian extensions of $\mathbb{Q}$.

We begin by fixing some notation. Let k be a real abelian extension of $\mathbb{Q}$ whose integer ring $\mathcal{O}$ is a unique factorization domain. Let M be an integer such that $k \subseteq \mathbb{Q}(\zeta_M)$ where ζ_M is a primitive M^{th} root of unity. Let D be the discriminant of the field $\mathbb{Q}(\zeta_M)$, so all primes of $\mathbb{Z}$ congruent to one modulo D split completely in $\mathcal{O}$. We will denote by $G = Gal(k/\mathbb{Q})$ the Galois group, and by N the norm from k to $\mathbb{Q}$. For $\alpha \in \mathcal{O}$, define the *powerful part of* α to be

$$\prod_{\mathfrak{p}^2|\alpha} \mathfrak{p}^{ord_{\mathfrak{p}}(\alpha)},$$

where $ord_{\mathfrak{p}}(\beta)$ is the exact power of $\mathfrak{p}$ dividing β. We denote by Φ_n the n^{th} cyclotomic polynomial. Now let $\alpha \in \mathcal{O}$ be any element such that one conjugate of α is larger than one in absolute value, and all other conjugates are smaller than one in absolute value. Put $\alpha^n - 1 = U_n V_n$, $\Phi_n(\alpha) = u_n v_n$ where V_n and v_n are the powerful parts of $\alpha^n - 1$ and $\Phi_n(\alpha)$ respectively. If p is a rational prime, define m_p to be the order of α modulo p, that is the least integer n such that $\alpha^n \equiv 1(p)$. Finally, for $X > 0$, define the set of *Wieferich primes up to* X *for* α to be

$$W(X) = \{\mathfrak{p} \in \mathcal{O} \text{ split prime with } p = N\mathfrak{p} \leq X \text{ and } \alpha^{p-1} \not\equiv 1 \ (mod\, \mathfrak{p}^2)\}.$$

We will use the symbol $a|b$ to mean *a divides b in the ring* $\mathcal{O}$, and $\mathfrak{p}$ will always be a prime (element) of the ring $\mathcal{O}$. We wish to prove that $|W(X)| \gg \log X$. Since $\alpha^{N\mathfrak{p}} - 1 \in \mathfrak{p}^2$ implies $\sigma\alpha^{N\sigma\mathfrak{p}-1} - 1 \in \sigma\mathfrak{p}^2$ for any $\sigma \in G$, we can assume that $|\alpha| \geq 1 \geq |\sigma(\alpha)|$ for all $\sigma \in G$. Also, since the norm of any split prime $\mathfrak{p}$ of $\mathcal{O}$ is a rational prime which can be assumed odd (with the possible exclusion from $W(X)$ of primes of $\mathcal{O}$ above two), $N(\mathfrak{p}) - 1$ can be assumed even, and thus we can replace α by $-\alpha$ if necessary to assume that $\alpha > 0$. Finally, the fact that for any positive integer k, $\alpha^{k(N\mathfrak{p}-1)} \not\equiv 1 \ (mod\, \mathfrak{p}^2)$ implies that $\alpha^{N\mathfrak{p}-1} \not\equiv 1 \ (mod\, \mathfrak{p}^2)$, we can replace α by any positive power of α. This allows us to assume that $\alpha > 2$, that all conjugates of α are positive, and that $\alpha - 1$ is not a unit, that is that $|N(\alpha - 1)| \geq 2$, since for any $\alpha \in \mathcal{O}$ with exactly one conjugate bigger than one in absolute value, $|N(\alpha^k - 1)|$ tends to infinity as k does. Note also that the condition $\alpha > 1$, $\sigma(\alpha) < 1$ for all $\sigma \in G$, $\sigma \neq Id$ implies that $N(\alpha^k - 1) \leq (\alpha^k - 1)$. We will assume these conditions on α without further comment.

Lemma 1 *Let n be a positive integer. Assume the prime $\mathfrak{p}$ is unramified in $\mathcal{O}$. If $\mathfrak{p}|U_n$ and $\mathfrak{p} \not| n$, then $n = m_p$, and $\alpha^{N(\mathfrak{p})-1} \not\equiv 1 \ (mod\, \mathfrak{p}^2)$ if and only if $\mathfrak{p}$ is a prime of degree one.*

Proof: Let $\mathfrak{p}$ be a prime dividing U_n, $\mathfrak{p} \not| n$. Since $\mathfrak{p}|U_n$, $\alpha^n \equiv 1 \ (mod \mathfrak{p})$. Now the polynomial $X^n - 1$ has distinct roots modulo $\mathfrak{p}$ (since $X^n - 1$ and nX^{n-1} are relatively prime modulo $\mathfrak{p}$), and thus the fact that $\Phi_n(\alpha) \equiv 0 \ (mod\, \mathfrak{p})$ implies that $\Phi_d(\alpha) \not\equiv 0 \ (mod\, \mathfrak{p})$ for any $d|p$, so $m_p = n$. Since $ord_{\mathfrak{p}} U_n = 1$, $\alpha^n \not\equiv 1 \ (mod\ \mathfrak{p}^2)$, so $\alpha^n = 1 + \beta\mathfrak{p}$ for some β a $\mathfrak{p}$-unit. Since $\alpha^{\phi(N\mathfrak{p})} \equiv 1 \ (mod\, \mathfrak{p})$, $n|\phi(N\mathfrak{p})$. Thus we can raise α^n to the power $\phi(N\mathfrak{p})/n$; this yields

$$\alpha^{\phi(N\mathfrak{p})} = (\alpha^n)^{\phi(N\mathfrak{p})/n} \equiv 1 + \frac{\beta\mathfrak{p}\phi(N\mathfrak{p})}{n} \quad (mod\, \mathfrak{p}^2).$$

Thus $\alpha^n \not\equiv 1(\mathfrak{p}^2)$ if and only if $\mathfrak{p} \not| \phi(N\mathfrak{p})$. If $\mathfrak{p}$ is degree one, $N\mathfrak{p} = p$ for some rational prime p, so $\mathfrak{p} \not| \phi(N\mathfrak{p}) = p - 1$, and if $\mathfrak{p}$ is not degree one, $Np = p^f$ for some $f > 1$ so $p|\phi(Np) = p^{f-1}(p-1)$.

With the hypotheses as in Lemma 1, we note that if $\mathfrak{p}$ is a prime with norm p^f for $f > 1$ and a rational prime p, the fact that $n|\phi(N\mathfrak{p}) = p^{f-1}(p-1)$ implies $n|(p-1)$ since $n \not| p$, so $p \equiv 1 \ (mod\ n)$. Thus, if $n \equiv 0 \ (mod\ D)$, p splits in k, so $\mathfrak{p}$ is of degree one, a contradiction. This gives the following

Lemma 2 *If $n \equiv 0 \ (mod\ D)$, all primes dividing U_n are of degree one.*

Lemma 3 $|W(X)| \geq |\{n \leq \log_\alpha(X) \ : \ N(U_n) > Dn \ and\, n \equiv 0 \ (mod\ D)\}|$

Proof: Let $n \equiv 0 \ (mod\ D)$ be such that $N(U_n) > Dn$. By lemma 2, $N(U_n)$ is square free, and thus we can choose $p|U_n$ prime to nD. Let $\mathfrak{p}_n$ be a prime of $\mathcal{O}$ above p. Applying lemma 1 gives that $m_{\mathfrak{p}_n} = n$ and $\alpha^{p-1} \not\equiv 1 \ (mod\, \mathfrak{p}_n^2)$. Since $n \leq \log_\alpha(X)$, $\alpha^n \leq X$, and

$$N(\mathfrak{p}_n) \leq |N(U_n)| \leq |N(U_n V_n)| \leq |N(\alpha^n - 1)| \leq (\alpha^n - 1) < \alpha^n \leq X.$$

The fact that $m_{p_n} = n$ means that different values of n give different primes, so the lemma is proved.

In order to apply lemma 3 to prove the infinitude of Wieferich primes for α, we must show that there exist an infinitude of natural numbers n satisfying the conditions implicit in lemma 3, that is $n \equiv 0 \ (mod\ D)$ for which $N(U_n) > D$. To do this, we first show that $N\Phi_n(\alpha)$ grows 'quickly' with n, and then that the uniform *abc*-conjecture for the field k implies that $N(V_n)$ grows 'slowly' with n. This is accomplished in the following two lemmas.

Lemma 4 *There exists a constant $c > 0$ such that $|N\Phi_n(\alpha)| > e^{c\phi(n)}$ for all n.*

Proof: For any $\sigma \in G$, $\sigma(\alpha)$ is real and positive, hence for any root of unity ζ, $|\sigma(\alpha - 1)| \leq |\sigma(\alpha - \zeta)|$ so we have

$$|N\Phi_n(\alpha)| = \prod_{\sigma\in G} |\Phi_n(\sigma\alpha)| \geq \prod_{\sigma\in G} |\sigma\alpha - 1|^{\phi(n)}.$$

Since $\alpha > 2$, $(\alpha - 1)^{\phi(n)} > e^{c'\phi(n)}$ for some $c' > 0$, and the other terms in the product on the right hand side are in absolute value greater than or equal to

$$\prod_{\substack{\sigma\in G\\ \sigma\neq Id}} |\sigma\alpha - 1| = c'' > 0$$

so that the entire product is greater than or equal to $c''e^{c'\phi(n)} \geq e^{c\phi(n)}$ for a third constant $c > 0$.

Lemma 5 *The abc-conjecture for the field k implies that $N(V_n) \ll \alpha^{\epsilon n}$ for any $\epsilon > 0$ (with the implied constant depending on α, k and ϵ).*

Proof: Because $U_nV_n = \Phi_n(\alpha)|(\alpha^n - 1) = u_nv_n$, $V_n|v_n$, so $N(V_n)|N(v_n)$, so it suffices to prove the estimate for $N(v_n)$. Applying the *abc*-conjecture for k to the equation $\alpha^n - u_nv_n - 1 = 0$ yields

$$H(\alpha^n, u_nv_n, 1) \ll G(\alpha^n, u_nv_n, 1)^{1+\epsilon},$$

with the constant depending on k and ϵ. The definitions of the height and the conductor show that

$$H(\alpha^n, u_nv_n, 1) = \prod_{\sigma\in G} \max\{|\sigma(\alpha^n)|, |\sigma(u_nv_n)|, 1\}^{1/d} \geq \alpha^{n/d}$$

$$G(\alpha^n, u_n v_n, 1) \ll \prod_{\mathfrak{q} | u_n v_n} N(\mathfrak{q})^{1/d} \leq N(u_n)^{1/d} N(v_n)^{1/2d},$$

since $\mathfrak{q} | v_n$ implies that $\mathfrak{q}^2 | v_n$. The implied constant in the second inequality depends only on α. We also know that $N(u_n v_n) = N(\alpha^n - 1) \leq \alpha^n - 1 < \alpha^n$, so $N(u_n) < \alpha^n / N(v_n)$. Therefore the *abc*-conjecture for k implies that

$$\alpha^{n/d} \ll \left(\frac{\alpha^{n/d}}{N(v_n)^{1/2d}} \right)^{1+\epsilon}.$$

Isolating $N(v_n)$ and replacing ϵ by $\epsilon/(2-\epsilon)$ gives the result.

Now these lemmas, taken together, suggest that we should look for values of n congruent to zero modulo D with large values of $\phi(n)$ relative to n (that is, with values of $\phi(n)/n$ as large as possible); the following lemma shows that there are 'many' such n.

Lemma 6 $\sum_{n \leq X} \frac{\phi(n)}{n} = \frac{6}{\pi^2} X + O(1)$.

Proof: We will show that both sides of the equation are equal to

$$\sum_{n \leq X} \frac{\mu(n)}{n} \left[\frac{X}{n} \right].$$

Theorem 3.11 of Apostol [2] states that for any arithmetical function f,

$$\sum_{n \leq X} f(n) \left[\frac{X}{n} \right] = \sum_{n \leq X} \sum_{d | n} f(d).$$

Applying this with $f(n) = \mu(n)/n$ and using the fact that $\phi(n) = \sum_{d|n} \mu(d) \frac{n}{d}$ gives

$$\sum_{n \leq X} \frac{\mu(n)}{n} \left[\frac{X}{n} \right] = \sum_{n \leq X} \frac{\phi(n)}{n}.$$

On the other hand, the fact that for $\Re(z) > 1$,

$$\frac{1}{\zeta(z)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n}$$

shows that

$$\sum_{n \leq X} \frac{\mu(n)}{n^2} = \frac{1}{\zeta(2)} - \sum_{n > X} \frac{\mu(n)}{n^2} = \frac{6}{\pi^2} - O\left(\frac{1}{X}\right),$$

since $\zeta(2) = \pi^2/6$. Therefore we have

$$\frac{X}{\zeta(2)} - O(1) = X \sum_{n \leq X} \frac{\mu(n)}{n^2} = \sum_{n \leq X} \frac{\mu(n)}{n} \left[\frac{X}{n}\right] + \sum_{n \leq X} \frac{\mu(n)}{n} \left(\frac{X}{n} - \left[\frac{X}{n}\right]\right),$$

and

$$\sum_{n \leq X} \frac{\mu(n)}{n} \left(\frac{X}{n} - \left[\frac{X}{n}\right]\right) \leq \sum_{n \leq X} \frac{\mu(n)}{n}$$

with the latter sum in absolute value less than or equal to one, by Theorem 3.13 of Apostol [2]. Thus, the lemma is proved.

Lemma 7 *For any positive integer A and any $\delta \in (0, \frac{6}{\pi^2}\frac{\phi(A)}{A})$,*

$$|\{n \leq X : \ \phi(n) \geq \delta n \ and \, n \equiv 0 \ (mod \ A)\}| \geq \frac{X}{A}\left(\frac{6}{\pi^2} - \delta \frac{A}{\phi(A)}\right) + O(1).$$

Proof: If $A = 1$, for $\delta \in (0, 6/\pi^2)$, we simply divide the sum from the lemma 6 into two parts:

$$\sum_{n \leq X, \, \phi(n) \geq \delta n} \frac{\phi(n)}{n} \leq |\{n \leq X : \ \phi(n) \geq \delta n\}|$$

$$\sum_{n \leq X, \, \phi(n) < \delta n} \frac{\phi(n)}{n} \leq \delta X.$$

Rearranging and applying lemma 6 gives the result for $A = 1$. For other values of A, we know that $\phi(nA) \geq \phi(n)\phi(A)$, so if $\phi(n) \geq \frac{\delta A}{\phi(A)} n$, $\phi(An) \geq \delta An$, and we have the result for arbitrary A.

Theorem 19 *Let k be a real abelian extension of $\mathbb{Q}$ whose integer ring $\mathcal{O}$ is a unique factorization domain, and let D be defined as above. Let $\alpha \in \mathcal{O}$ have all but one conjugate less than one in absolute value. Defining $W(X) = W_\alpha(X)$ as above, the uniform abc-conjecture for k implies that $|W(X)| \gg \log(X)$.*

Proof: By lemma 3, it is enough to prove that

$$|\{n \leq \log_\alpha(X) \ : \ N(U_n) > Dn \text{ and } n \equiv 0 \ (mod\ D)\}| \gg \log X.$$

$N(U_nV_n) = N\Phi_n(\alpha)$, so by lemma 4 and lemma 5 there exist constants c and $c' = c'_\epsilon$ such that

$$N(U_n) \geq \frac{e^{c\phi(n)}}{c'\alpha^{\epsilon n}},$$

so we see that any n congruent to zero modulo D for which $c\phi(n) \geq c_2 + \log n + \epsilon n \log\alpha$ ($c_2 = \log Dc'$) will provide a Wieferich prime for α. Pick δ small enough for lemma 7 to apply, and put $\epsilon = \delta c/2\log\alpha$. Then $\phi(n) \geq \delta n$ implies that

$$c\phi(n) - c_2 - \epsilon n \log\alpha - \log n > \delta cn - c_2 - \frac{\delta cn}{2} - \log n$$

which will be greater than zero if n is sufficiently large, say if $n \geq n_0$. Thus

$$\begin{aligned} |W(X)| &\geq |\{n_0 \leq n \leq \log_\alpha(X) : n \equiv 0 \ (mod\ D) \text{ and } \phi(n) \geq \delta n\} \\ &\geq \frac{\log_\alpha(X)}{D}\left(\frac{6}{\pi^2} - \frac{D\delta}{\phi(D)}\right) + O(1) - n_0, \end{aligned}$$

which proves the theorem.

4.2 Connection with the Euclidean Algorithm

The original motivation in studying this generalization of Silverman's work was an attempt to prove that the uniform *abc*-conjecture implies that the rings of certain number fields which are unique factorization domains, are also Euclidean domains. This problem was first suggested by Hasse [12]. Weinberger [33] proved under the assumption of the generalized Riemann hypothesis for certain Dedekind ζ-functions that the integer ring of a number field which is not imaginary quadratic is a Euclidean domain if and only if it is a unique factorization domain. In [4], Clark and R. Murty prove unconditionally that if R is the integer ring of a totally real Galois extension of $\mathbb{Q}$ of degree n, then R is a Euclidean domain if there exist $m = |n-4|+1$ split prime ideals of R, $\mathfrak{p}_1, \cdots, \mathfrak{p}_m$ such that for every $(k_1, \cdots, k_m) \in \mathbb{N}^m$, the natural projection map

$$U \to \left(\frac{R}{\mathfrak{p}_1^{k_1} \cdots \mathfrak{p}_m^{k_m}}\right)^*$$

is a surjection, where U is the unit group of R and $(A)^*$ denotes the unit group of the ring A. They also prove that for this condition to hold for every m-tuple $(k_1, \cdots, k_m) \in \mathbb{N}^m$, it is sufficient that it hold for $(2, 2, \cdots, 2)$. The simplest examples of real fields whose integer rings are unique factorization domains but which aren't known to be Euclidean domains are integer rings of real quadratic fields. For example, $\mathbb{Z}[\sqrt{14}]$ is a unique factorizaton domain, but the norm function does not provide an Euclidean algorithm in this ring; it is not known whether or not the ring is in fact a Euclidean domain. However, the theorem of Clark and Murty cannot be applied when R is the integer ring of a real quadratic field, since in this case it is necessary to find three primes $\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3$ such that

$$U \twoheadrightarrow \left(\frac{R}{\mathfrak{p}_1^2\mathfrak{p}_2^2\mathfrak{p}_3^2}\right)^*;$$

and $U = \{\pm 1\}\times < u >$, where $< u >$ is the (infinite cyclic) group generated by the fundamental unit u. Inasmuch as the order of each group $(R/\mathfrak{p}_i^2)^*$ is even,

$$\left(\frac{R}{\mathfrak{p}_1^2\mathfrak{p}_2^2\mathfrak{p}_3^2}\right)^* \cong \left(\frac{R}{\mathfrak{p}_1}\right)^* \times \left(\frac{R}{\mathfrak{p}_2}\right)^* \times \left(\frac{R}{\mathfrak{p}_3}\right)^*$$

cannot contain a cyclic group of index less than four, and thus the image of $\{\pm 1\}\times < u >$ must be of index at least two. However, the theorem should be applicable to totally real Galois extensions of degree at least three over $\mathbb{Q}$, since for these fields the number m defined above will be less than or equal the number of multiplicatively independent elements of U. Let R be any ring whose field of fractions k is a totally real Galois extension of $\mathbb{Q}$ of degree $n \geq 3$. If $\mathfrak{p}$ is any split prime of R and $u \in U \setminus \{\pm 1\}$ then $\pm < u > \twoheadrightarrow R/\mathfrak{p}^2$

if and only if the following two conditions are met:

$$u^{N\mathfrak{p}-1} \not\equiv 1 \ (mod\, \mathfrak{p}^2) \tag{4.1}$$

$$\pm < u > \twoheadrightarrow \frac{R}{\mathfrak{p}} \tag{4.2}$$

The following lemma says that there are many elements of the unit group $U = R^*$ to which the result of section one can be applied.

Lemma 1 *If R is as above then for any $\sigma \in G = Gal(k/\mathbb{Q})$ there exist $n-1$ multiplicatively independent units $u_1, \cdots, u_{n-1}$ of R with the property that $|\sigma(u_i)| > 1$ and $|\tau(u_i)| \leq 1$ for all $\tau \in G$, $\tau \neq \sigma$.*

Proof: Say $G = \{\sigma_1, \cdots, \sigma_n\}$, $\sigma_1 = Id$. It is clearly sufficient to prove the lemma for $\sigma = Id$. The Dirichlet unit theorem (see [16]) states that the map $\phi : U \to \mathbb{R}^{n-1}$ given by

$$u \mapsto (\log |u|, \log |\sigma_2(u)|, \cdots, \log |\sigma_n(u)|)$$

sends U onto a full lattice of $\mathbb{R}^{n-1}$ (where $\mathbb{R}^{n-1}$ is identified with the hyperplane of $\mathbb{R}^n$ given by the equation $x_1 + \cdots + x_n = 0$). The condition that $|u| > 1$, $\sigma_i(u) < 1$ for $i \neq 1$ is equivalent to having $\phi(u)$ lie in the '2^{n-1}-tant' (quadrant, octant, ...) W consisting of points with first coordinate positive and all other coordinates negative. This region contains $n-1$ linearly independent vectors of $\phi(U)$ since otherwise $W \bigcap \phi(U)$ would be contained in a hyperplane Π of $\mathbb{R}^{n-1}$, but there exists an element $x \in \phi(U) \setminus \Pi$, and by adding x to a suitable element of $\phi(U) \bigcap W$ we can find an element of $\phi(U) \bigcap W$ not contained in Π.

Thus, the result of section one applies to $u_1, \cdots u_n$, and shows that the *abc*-conjecture for k implies that there exist infinitely many split primes such that $u = u_i$ satisfies (4.1) for $i = 1, \cdots n-1$. Furthermore, we can choose these elements not to be perfect squares - if (say) $u_1 = v^r$ for some even integer r, and v is not a perfect power in R, then since u_1 has exactly one conjugate greater than one in absolute value, so does v and hence we can replace u_1 by v. It has been conjectured by E. Artin that any integer not equal to ± 1 or a perfect square is a primitive root modulo a positive density of primes (see [1, introduction] for details) and analogs of this conjecture in the number field case have been proven by H.W. Lenstra Jr. ([17]), modulo generalized Riemann hypotheses for certain Dedekind ζ-functions. R. Gupta and M.R. Murty, in [11] prove that all but at most twelve rational primes are primitive roots modulo infinitely many primes, and W. Narkiewicz uses similar methods and an improvement of Heath-Brown ([13]) in [24] to prove a similar result for certain Abelian extensions of $\mathbb{Q}$. Specifically, Narkiewicz proves that in any Abelian extension of $\mathbb{Q}$ satisfying a certain technical condition, there are at most two (non-square) multiplicatively independent elements for which Artin's conjecture fails. Thus, in a real Abelian extension k of degree greater than or equal to four, there are (at least) as many multiplicatively independent units which are primitive roots modulo infinitely many prime ideals as the number m of primes required for the application of Clark and Murty's result, so assuming abc there are infinitely many primes satisfying (4.1) and infinitely many primes satisfying (4.2). Unfortunately, however, we require primes satisfying both of these conditions simultaneously, and this appears to be difficult–Narkiewicz' work shows primitivity modulo primes p such that $p-1$ has the form $2q_1q_2$ or $2q_1$ for q_1, q_2 prime; and there is no (obvious) way to centre out primes of this form using Silverman's method. However, it seems reasonable that there should exist only finitely many non-Weiferich primes for a given field element given the paucity of primes p such that $2^{p-1} \equiv 1 \ (mod\, p)$ (although the method used in the previous section is inadequate for proving that abc implies this result), and this, together with Narkiewicz' and Clark and Murty's results would imply that for $\mathcal{O}$ the integer ring of a number field k which is not quadratic, $\mathcal{O}$ is Euclidean if and only if $\mathcal{O}$ is a p.i.d.

Bibliography

[1] E. Artin, ed. S. Lang and J. Tate Collected Works Reading, Mass., Addison-Wesley Publishing Co. Inc. 1965

[2] T. Apostol, Introduction to Analytic Number Theory New York, Springer-Verlag 1976

[3] J. Cassels, Rational Quadratic Forms London, Academic Press 1978

[4] D. Clark and R. Murty, *The Euclidian Algorithm in Galois Extensions of* $\mathbb{Q}$ J. Reine und Angew Math., **459** (1995), 151-162

[5] H. Davenport, Multiplicative Number Theory (2nd Edition) New York, Springer-Verlag 1980

[6] P. Erdös, *Problems and Results on Consecutive Integers*, Eureka **38** (1975/76)

[7] G. Frey, *Links between solutions of* $a - b = c$ *and Elliptic Curves* in Number Theory, Ulm 1987 (Lecture Notes in Mathematics 1380) New York, Springer-Verlag 1989

[8] A. Granville, *Problems Related to Fermat's Last Theorem* in Number Theory, Banff 1988 R.A. Mollin (ed.), Berlin, de Gruyter 1990

[9] A. Granville, *The First Case of Fermat's Last Theorem is True for all Prime Exponents up to 714, 591, 416, 091, 389* Trans. Amer. Math. Soc **306** (1988) 329-359

[10] A. Granville and H.M. Stark, *ABC Implies No 'Siegel Zeroes'* to appear.

[11] R. Gupta and M.R. Murty, *A Remark on Artin's Conjecture* Invent. Math. **78** (1984) 127-130

[12] H. Hasse, *Zerlegung in Primelemente oder in Primhauptideale in Integritätsbereichen* J. Reine und Angew. Math., **159** (1928) 3-12

[13] D.R. Heath-Brown *Artin's Conjecture for Primitive Roots* Quart. J. Math. Oxford (2) **37** (1986) 27-38

[14] I. Katznelson, An Introduction to Harmonic Analysis New York, Dover Publications Inc. 1976

[15] S. Lang, Elliptic Functions Reading, Mass., Addison-Wesley Pub. Co. 1973

[16] S. Lang, Algebraic Number Theory New York, Springer-Verlag 1986

[17] H.W. Lenstra, Jr., *On Artin's Conjecture and Euclid's Algorithm in Global Fields* Invent. Math. **42** (1977) 201-224

[18] K. Mahler, *On Hecke's Theorem on the Real Zeroes of L-functions and the Class Numbers of Quadratic Fields* J. London Math Soc. **9** (1934) 298-302

[19] L. Mai and R. Murty *The Phragmen-Lindelöf Theorem and Modular Elliptic Curves* Contemp. Math. **166** (1994) 335-340

[20] R. C. Mason, Diophantine Equations over Function Fields Cambridge, Cambridge University Press 1984

[21] D.W. Masser, *Open Problems* in Proc. Symp. Analytic Number Th., W.W.L. Chen (ed.), London, Imperial College 1985

[22] R.A. Mollin and R.G. Walsh, *A Note on Powerful Numbers, Quadratic Fields and the Pellian*, C.R. Math. Acad. Sci. Canada **8** (1986), pp. 109-114

[23] L.J. Mordell, *On the Riemann Hypothesis and Imaginary Quadratic Fields with a Given Class Number* J. London Math. Soc. **9** (1934) 289-298

[24] W. Narkiewicz *A note on Artin's conjecture in algebraic number fields* J. Reine und Angew. Math **381** (1987) 110-115

[25] Marius Overholt, *The Diophantine equation* $n! + 1 = m^2$ Bull. London Math. Soc. **25** (1993) 104

[26] A. Borel, et. al. Seminar on Complex Multiplication (Lecture Notes in Mathematics 21) Berlin, Springer-Verlag 1966

[27] J. Silverman, The Arithmetic of Elliptic Curves New York, Springer-Verlag 1986

[28] J. Silverman, *Wieferich's Criterion and the abc-Conjecture* J. Number Theory **30** (1988) 226-237

[29] C.L. Stewart and R. Tijdeman, *On the Oesterlé-Masser conjecture* Mh. Math **102** (1986) 251-257

[30] C.L. Stewart and K. Yu, *On the abc conjecture* Math. Ann. **291** (1991) 225-230

[31] P. Vojta Diophantine Approximation and Value Distribution Theory (Lecture Notes in Mathematics 1279) New York, Springer-Verlag 1987

[32] H. Weber, Lehrbuch der Algebren vol. III. New York, Chelsea Publishing Co.

[33] P.J. Weinberger, *Euclidean rings of algebraic integers* in Proc. Symp. Pure Mathematics, vol. XXIV, Providence, R.I., American Mathematical Society 1973

[34] A. Wieferich, *Zum Letzen Fermat'schen Theorem* J. Reine und Angew Math., **136** (1909) 293-302