Markos Karameris

Applications of the Modularity Theorem to Diophantine Equations

Master's Thesis

University of Athens, Department of Mathematics

Athens September 15, 2021

Advisor: Aristides Kontogeorgis

Master Thesis Committee

Mihalis Maliakas

Aristides Kontogeorgis

Ioannis Dokas

To my mentor Angelos for his constant support,

Contents

Εισαγωγή **ix**

Introduction **xi**

1 Elliptic Curves **1**
1.1 Basic Definitions 1
1.2 Weierstrass Equations and the group law 2
1.3 Isogenies 3
1.4 Tate Module 4
1.5 Weil Pairing 5
1.6 Local fields and reduction properties 6
1.7 L-functions of elliptic curves 8
1.8 Elliptic Curve Representations 9

2 Modular Forms **13**
2.1 Basic Definitions 13
2.2 Modular curves and Dimension 14
2.3 Hecke Operators 15
2.3a Double coset operators 16
2.3b Diamond and T_p operators 17
2.3c Petersson Inner Product 18
2.3d Oldforms and Newforms 19
2.4 L-functions associated to modular curves 20
2.5 Jacobians 21
2.5a Maps between Jacobians 21
2.5b Jacobians of Modular Curves 22
2.6 Algebraic Eigenvalues 22
2.7 Abelian Varieties and Newforms 23
2.8 From Geometry to Algebra 24
2.9 Representations and Modularity 26

3 The Modular Approach 29
3.1 Level Lowering 29
3.2 The Modular Approach 31
3.2a Fermat's Last Theorem 31
3.2b Generalizing the Method 31
3.3 Applications to Diophantine Equations 32

Literature 35

Εισαγωγή

Ο συνηθέστερος τρόπος να επιδείξει κανείς τις εφαρμογές της Modular μεθόδου στις Διοφαντικές εξισώσεις είναι αναμφίβολα το Τελευταίο Θεώρημα του Fermat(FLT). Η απόδειξη του από τον Wiles αποτελεί ένα από τα ορόσημα των σύγχρονων μαθηματικών και αποτελεί πράγματι ένα ταξίδι που περνάει από πολλούς κλάδους, συνδέοντας αναλυτικά αντικείμενα (δομοστοιχειωτές μορφές) με αλγεβρικά (Galois αναπαραστάσεις) και γεωμετρικά αντικείμενα (ελλειπτικές καμπύλες). Μια δομοστοιχειωτή μορφή περιγράφεται συχνά ως «μια μιγαδική συνάρτηση με εξέχουσα συμμετρία και κάποιες επιπλέον συνθήκες ολομορφίας». Κάθε δομοστοιχειωτή μορφή έχει δύο αριθμούς που συνδέονται με αυτήν, το «επίπεδο» N και το «βάρος» k. Από την άλλη μεριά αν θεωρήσουμε κυβικές εξισώσεις της μορφής $y^2 = 4x^3 - g_2x - g_3, g_2, g_3 \in \mathbb{Z}$ με $\Delta = g_2^3 - 27g_3^2 \neq 0$ τότε αυτές ορίζουν ομαλές καμπύλες που ονομάζουμε «ελλειπτικές καμπύλες». Το ανάλογο του επιπέδου σε αυτή τη περίπτωση είναι μια ποσότητα που ονομάζουμε «οδηγό» της καμπύλης. Η σύνδεση αυτών των δύο αντικειμένων είναι αποτέλεσμα του Modularity Theorem που ήταν παλαιότερα γνωστό ως εικασία Taniyama–Shimura–Weil. Αυτή η εικασία υπήρχε ανεξάρτητα από το FLT από τη δεκαετία του 1960. Η ακριβής διατύπωση είναι:

«Κάθε ελλειπτική καμπύλη πάνω από τους ρητούς είναι modular»

Το δεύτερο βήμα προς την απόδειξη έγινε περίπου είκοσι χρόνια αργότερα όταν ο Gerhard Frey είκασε ότι αν επιτρέψουμε μια μη τετριμμένη υποθετική λύση στην εξίσωση $x^n + y^n = z^n$, τότε η ελλειπτική καμπύλη $y^2 = x(x - a^n)(x - b^n)$ δεν θα είναι modular. Μετά από μία μερική απόδειξη του Serre, το τελικό χτύπημα δώθηκε από τον Ribet με το θεώρημα που έμεινε γνωστό ως «Ribet's Level Lowering theorem». Η ισχύς της μεθόδου έγκειται στο ότι σε μια καμπύλη όπως του Frey με $\Delta_E = 2^{-8}(abc)^n$ και οδηγό $N_E = \mathrm{rad}(abc)$ μπορούμε να αντιστοιχίσουμε μια συγκεκριμένου τύπου δομοστοιχειωτή μορφή γνωστή ως newform με επίπεδο N_f που δεν εξαρτάται από τα a, b, c. Στην περίπτωση του FLT, $N_f = 2$ και έχουμε άτοπο διότι δεν υπάρχουν newforms με βάρος 2 και επίπεδο 2.

Η Modular μέθοδος είναι μια γενίκευση αυτού του αποτελέσματος στην οποία αντιστοιχίζουμε μια Frey καμπύλη στη δεδομένη διοφαντική εξίσωση, υπολογίζουμε τις δυνατές newforms και είτε τις απορρίπτουμε είτε τις χρησιμοποιούμε για να φράξουμε κάποια από τις άγνωστες παραμέτρους μας.

Το πρώτο κεφάλαιο αφορά τις ελλειπτικές καμπύλες. Ξεκινάμε με βασικούς ορισμούς, το προσθετικό νόμο, την εξίσωση Weierstrass και επιγραμματικά αναφέρουμε τις ισογένιες. Συνεχίζουμε με τον ορισμό του προτύπου του Tate και την αντιστοίχηση του Weil ώστε να μιλήσουμε για $\ell-$αδικές αναπαραστάσεις ελλειπτικών καμπυλών και τη σύνδεση τους με $L-$συναρτήσεις. Κάνουμε ακόμη λόγο για

αναγωγές ελλειπτικών καμπυλών και για ελλειπτικές καμπύλες πάνω από τοπικά σώματα.

Στο δεύτερο κεφάλαιο έρχονται στο προσκήνιο οι δομοστοιχειακές μορφές. Ξεκινάμε πάλι με τους βασικούς ορισμούς και επεξηγούμε πως οι δομοστοιχειακές μορφές συγκεκριμένου επιπέδου και βάρους αποτελούν πεπερασμένης διάστασης διανυσματικό χώρο. Στη συνέχεια κάνουμε λόγο για τελεστές Hecke και τον χώρο των newforms $\mathcal{S}_k(\Gamma_1(N))^{new}$ που είναι το κυρίαρχο αντικείμενο που θα μας απασχολήσει. Περνάμε κατόπιν στις Ιακωβιανές που προέρχονται από δομοστοιχειακές καμπύλες και εξετάζουμε πως αυτές προβάλλονται ισογενώς σε ευθύ άθροισμα αβελιανών πολλαπλοτήτων A_f από newforms βάρους 2. Η «γέφυρα» μας από την γεωμετρία και το $\mathbb{C}$, στο $\mathbb{Q}$ και τη θεωρία αριθμών είναι η σχέση Eichler-Shimura και η αλγεβρο-γεωμετρική προσέγγιση των ανηγμένων καμπυλών μας. Όλα αυτά τελικά θα μας επιτρέψουν να συσχετίσουμε σε μια newform f μια Galois αναπαράσταση ως δράση της $G_{\mathbb{Q}}$ στο Tate πρότυπο $T_\ell(A_f)$.

Το τελευταίο κεφάλαιο είναι αφιερωμένο στην Modular μάθοδο και τις εφαρμογές της. Θα χρησιμοποιήσουμε το Modularity Theorem από τη σκοπιά των Galois αναπαραστάσεων για να σκιαγραφήσουμε τη μέθοδο και να αναδείξουμε την αποτελεσματικότητα της. Αυτό επιτυγχάνεται με συγκεκριμένα παραδείγματα από την εξίσωση $x^2 + d^2 = 2y^n$ με τη χρήση κώδικα σε SageMath για το υπολογιστικό σκέλος.

Αθήνα Σεπτέμβριος 2021.

Introduction

One cannot start an introductory discussion on the Modular Approach without mentioning Fermat's Last Theorem (FLT). Wile's proof of the FLT is one of the most celebrated results of our time and serves indeed as an excellent starting point for what is essentially a journey around all the realms of mathematics, linking analytic objects (modular forms) to algebraic objects (Galois representations) and through that correspondance to geometric objects (elliptic curves). A modular form is often described as "a complex function with exceeding symmetry satisfying certain holomorphy conditions". Each modular form has two numbers associated with it: a level N and a weight k. On the other hand if we consider cubic equations of the form $y^2 = 4x^3 - g_2x - g_3, g_2, g_3 \in \mathbb{Z}$ with $\Delta = g_2^3 - 27g_3^2 \neq 0$, then these equations define smooth curves which we call "elliptic curves". The analogue of the level in this case is a quantity called the "conductor" of the curve. The connection between these objects is the result of the famous Modularity Theorem previously known as the Taniyama–Shimura–Weil conjecture. This conjecture existed independently from FLT from the 1960s. The exact conjecture is:

"Every rational elliptic curve is modular."

The second step towards the proof happened 2 decades later when Gerhard Frey conjectured that if we allow a hypothetical solution non-trivial of the Fermat equation $a^n + b^n = c^n$ to exist, then the elliptic curve with equation $y^2 = x(x - a^n)(x + b^n)$ will not be modular. After a partial proof by Serre, the last part known as the "ϵ-conjecture" was proved by Ribet by a method known as "Level Lowering". The strength of the method lies in the fact that given our Frey curve with discriminant $\Delta_E = 2^{-8}(abc)^n$ and conductor $N_E = \mathrm{rad}(abc)$ it attaches to it a specific type of modular form f called a "newform" with level N_f independent of the unkowns a, b, c. In the case of FLT $N_f = 2$ and the contradiction is immediate since it turns out there are no newforms of weight 2 and level 2.

The Modular Approach is a generalization of this method where one finds a way to attach a Frey curve to a Diophantine equation, calculates all the possible newforms and proceeds to eliminate them or use them in clever ways to bound our unknown parameters.

Chapter 1 is devoted to the theory of elliptic curves. We begin with the definitions, discuss the group law, the Weierstrass equation and briefly touch on isogenies. Afterwards we define the Tate module and the Weil pairing so we can talk about the ℓ-adic Galois representations associated to elliptic curves and discuss the connection with $L-$functions of elliptic curves. We also briefly address elliptic curves over local fields and their reduction properties.

Chapter 2 is where we introduce modular forms. Again we start with the definitions and establish that modular forms of a certain level make up a finite dimensional vector space. We also discuss Hecke operators and the space of newforms $S_k(\Gamma_1(N))^{new}$ which is the kind of forms we are primarily interested in. Next we will briefly look at Jacobians of modular curves and how these decompose into a direct sum of abelian varieties A_f of weight 2 eigenforms. Then our aim is to shift from complex numbers to $\mathbb{Q}$ and view the curves we established from the perspective of algebraic geometry using the Eichler-Shimura relation. Finally this will allow us to associate to an eigenform f a Galois representation as the action of $G_{\mathbb{Q}}$ on the Tate module $T_\ell(A_f)$.

The last chapter is specifically dedicated to the Modular Approach and it's applications. We will use the Galois representation form of the Modularity Theorem to provide an overview of the method and illustrate it's effectiveness on specific instances of the diophantine equation $x^2 + d^2 = 2y^n$ backed by SageMath code.

Athens September 2021.

Chapter 1

Elliptic Curves

1.1 Basic Definitions

Elliptic curves are at the heart of the Modular Approach as we will see in Section 3. So the first natural question is what is an elliptic curve? Suppose k is a field, then:

Definition 1.1.1. *An elliptic curve over k is a smooth, projective plane curve of genus 1 with an additional k-rational point O. This distinguished point is also referred to as the "base point" or the "point at infinity" of the curve.*

Let us break down the definition a bit more: A projective plane curve C_f/k is a homogeneous polynomial $f(x, y, z)$ with coefficients in k. If $K \supset k$ then the $k-$rational points of C_f are the set $C_f(K) = \{(x : y : z) \in \mathbb{P}^2(K) : f(x, y, z) = 0\}$. By Jacobian criteria, a point P is singular if $\frac{\partial f}{\partial x}|_P = \frac{\partial f}{\partial y}|_P = \frac{\partial f}{\partial z}|_P = 0$. A smooth curve is one that has no singular points. Let E/k denote an elliptic curve over k. A very important tool that will also come into play later is the divisor group of the curve. The divisor group is the free abelian group generated by the points of E. In more detail:

Definition 1.1.2. *We denote* $\mathrm{Div}(E)$ *the divisor group of the elliptic curve E. The elements of* $\mathrm{Div}(E)$ *are the sums $D = \sum_{P \in E} n_P P$ where $n_P \in \mathbb{Z}$ and $n_P = 0$ for all but finitely many P. The degree of the divisor is* $\deg(D) = \sum_{P \in E} n_P$. *The elements of degree 0 form a subgroup* $\mathrm{Div}^0(E)$ *called the "group of zero divisors of E".*

The divisor group is very important as it allows us to track the zeroes and poles of rational functions over our curve. In particular if $k(E)$ is the function field of E/k as a projective variety then if $f \in k(E)$ we define $div(f) = \sum_{P \in E} ord_P P$. This leads us to consider the Picard group of the curve. We will always have $\deg(div(f)) = 0$ (our functions have the same number of zeroes and poles counting multiplicity) and thus the principal divisors are always in $\mathrm{Div}^0(E)$.

Definition 1.1.3. *A divisor D is called "principal" if $D = div(f)$ for some $f \in k(E)$. We also define an equivalence relation $D_1 \sim D_2 \iff D_1 - D_2$ is principal. The Picard group is then the quotient group* $\mathrm{Pic}(E) = \mathrm{Div}(E)/\sim$ *and similarly the 0-Picard group* $\mathrm{Pic}(E) = \mathrm{Div}^0(E)/\sim$

The motivation for the above definition will become clear once we discuss the group law of an elliptic curve.
One more useful notion is the vector space $\mathcal{L}(D) = \{f \in k(E) : div(f)+D \geq 0\}$, where the divisor $D \geq 0 \iff D = \sum n_P P$ and every $n_P \geq 0$. If we denote the corresponding dimension by $\ell(D)$ then we have that:

Lemma 1.1.4. *For every* $D \in \mathrm{Div}(E) : \deg(D) \geq 1 \implies \ell(D) = \deg(D)$

Proof. The proof is an easy application of Corollary 5.5 [3.3]. □

1.2 Weierstrass Equations and the group law

The most common way to describe an elliptic curve is through it's Weierstrass equation, that is the polynomial it satisfies as a plane curve in dehomogenized form.

Theorem 1.2.1. *Let* E/k *be an elliptic curve, then there exist* $a_1, a_2, a_3, a_4, a_6 \in k$ *and (coordinate) functions* $x, y \in k(E)$ *such that* $\phi : E \to \mathbb{P}^2(k)$ *with* $\phi : P \to [x(P), y(P), 1]$ *is an isomorphism between* E *and* $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ *(1) and* $\phi(O) = [0, 1, 0]$. *Moreover if there exist two such equation these are equivalent under a change of variables* $x = u^2x' + r, y = u^3y' + su^2x' + t$. *Conversely any smooth cubic of this form corresponds to an elliptic curve.*

Proof. Notice that $\ell(6O) = 6$ and a basis of $\mathcal{L}(6O)$ consists of the 7 functions $1, x, y, xy, x^2, y^2, x^3, y^2$ which implies they are linearly dependent over k which connects x, y with an equality of the form (1). For the complete detailed proof see Proposition 3.1 [3.3]. □

Since we have described elliptic curves via their corresponding Weierstrass equations we now proceed to look at some important invariants based on this equation. The most important ones are the j invariant and the discriminant Δ of the curve. Using the same notation for the a_i as above we also define the quantities:

$$b_2 = a_2^2+4a_4,\ b_4 = 2a_4+a_1a_3,\ b_6 = a_3^2+4a_6,\ c_4 = b_2^2-24b_4,\ c_6 = -b_2^3+36b_2b_4-216b_4$$

And using those we define:

$$\Delta = (c_4^3 - c_6^2)/1728,\ \ j = c_4^3/\Delta$$

The discriminant is an extremely useful invariant as it tells us exactly when our equation describes a non singular curve. The j invariant on the other hand tells us when two elliptic curves are isomorphic. The situation is summarized in the following Proposition:

Proposition 1.2.2. *A curve given by a Weierstrass equation:*

(i) *is nonsingular if* $\Delta \neq 0$

(ii) *has a nodal singularity if* $\Delta = 0$ *and* $c_4 \neq 0$

(iii) *has a cusp if* $\Delta = c_4 = 0$

Furthermore two elliptic curves E_1, E_2 *both over* k *are isomorphic if and only if* $j(E_1) = j(E_2)$.

Proof. See Proposition 1.4 [3.3]. □

The major interest in elliptic curves emerged from a really strong property they possess: elliptic curves are abelian varieties meaning that the points on the curve form an abelian group under some operation. In particular there is a way to add any two points on the curve and get a third point. The construction is the following: take two points $P, Q \in E$ and find the third point R where the curve intersects the line formed by P and Q. Reflect the point R around the x-axis and you get the point $P + Q$. The point O is the identity of our group and in the case $P = Q$ we can consider the tangent and proceed similarly. With this construction it is obvious that $P + Q + R = 0$ as opposite points sum to O.

This group might seem familiar with something we already defined and indeed it is! There is an isomorphism $\kappa : E \to \mathrm{Pic}^0(E)$ identifying the 0 Picard group with the points on the curve. The construction of κ follows from the following proposition.

Proposition 1.2.3. *Let E be an elliptic curve then:*

(i) *For every $D \in \mathrm{Div}^0(E)$ there exists a unique $P \in E : D \sim P - O$ and we define $\sigma : \mathrm{Div}^0(E) \to E$ with $\sigma(D) = P$,*

(ii) *σ is 1-1 in $\mathrm{Pic}^0(E)$,*

(iii) *There exists an isomorphism $\kappa : E \to \mathrm{Pic}^0(E)$.*

Proof. (i) Observe that $\ell(D + O) = \deg(D + O) = 1$ and thus $\exists f \in k(E) : -D - O \leq div(f)$ with $div(f) = 0 \implies div(f) = P - D - O \implies D \sim P - O$

(ii) $\sigma(D_1) = \sigma(D_2) \iff P_1 - O = P2 - O \iff P_1 = P_2$

(iii) Simply set $\kappa = \sigma^{-1}$

□

We sum up this section with a quantity defined on an elliptic curve which is the invariant differential

$$\omega = \frac{dx}{2y + a_1 x + a_3}$$

written in the usual Weierstrass form notation. It is a straightforward calculation to show that this differential is invariant under the group action $\tau_Q(P) = P + Q$.

1.3 Isogenies

We have so far discussed the basic notions of elliptic curves and concluded that a curve can be naturally identified with it's Picard group as a group of points. The next step is to discuss the morphisms between these curves and specifically those that preserve the group structure. Let ϕ be a morphism of elliptic curves $\phi : E_1 \to E_2$ and denote by $\phi^* : k(E_1) \to k(E_1)$ the pullback map induced on function fields by sending $f \to f \circ \phi$.

Definition 1.3.1. *An isogeny between two elliptic curves E_1/k and E_2/k is a morphism of curves $\phi : E_1 \to E_2$ such that $\phi(O) = O$. The degree of the isogeny is the degree of the extension $[K(E_1) : \phi^*(K(E_1))]$ with $K = \bar{k}$. We denote $\mathrm{Hom}_k(E_1, E_2)$ the additive group of isogenies between E_1 and E_2 over k.*

An isogeny is actually a group homomorphism of elliptic curves. This can be seen by noting that the pushforward map between the Picard groups is either trivial or a homomorphism. This is immediate from II.3.7 [3.3].
The first useful fact on isogenies is that their kernel is finite.

Proposition 1.3.2. *Let $\phi : E_1 \to E_2$ be an isogeny, then $|ker(\phi)| = \deg_s(\phi)$ which is equal to the degree of the seperable part of the extension $[K(E_1) : \phi^*(K(E_1))]$. In fact $|ker(\phi)| = |\phi^{-1}(Q)|, \forall Q \in E$ and in particular $\ker(\phi)$ is finite.*

Proof. See Theorem 4.10, p. 72 [3.3]. □

Another important fact on isogenies is that the notion of "being isogenous to" is an equivalence relation. This is established by the existence of the dual isogeny.

Theorem 1.3.3. *Let $\phi : E_1 \to E_2$ be an isogeny of elliptic curves over k. Then there exists an isogeny $\hat{\phi} : E_2 \to E_1$ such that $\hat{\phi} \circ \phi = [m]_{E_1}$ and $\phi \circ \hat{\phi} = [m]_{E_2}$ where m is the degree of ϕ.*

One can show that if $\phi, \psi \in \mathrm{Hom}(E_1, E_2)$ then $\widehat{\phi + \psi} = \hat{\phi} + \hat{\psi}$. As we will later see there are many properties that are shared between curves in the same isogeny class.
There is an interesting class of isogenies $\phi : E \to E$. Those are the multiplication by m maps on the curve which are defined as $[m] : E \to E$ with $[m]P = \underbrace{P + ... + P}_{m \text{ times}} = mP$. This is trivially seen to be an isogeny. Below we will discuss the kernel of these isogenies and we will use them in constructing the Tate module. We denote $E[m] = \{P \in E : mP = 0\}$ that is the kernel of the $[m]$ isogeny.

Proposition 1.3.4. *Set $p = \mathrm{char}(k), p \nmid m$ then $E[m] = \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$. Otherwise:*

(i) $E[p^e] = O$ *or*

(ii) $E[p^e] = \mathbb{Z}/p^e\mathbb{Z}$

1.4 Tate Module

Let $K = \bar{k}$ and consider the Galois group $G_{K/k}$. As we mentioned in the introduction, our goal is to aquire a representation of $G_{K/k}$ that is related to our elliptic curve. We can start by observing that for every $P \in E[m]$ and $\sigma \in G_{K/k}$ we have $mP^\sigma = (mP)^\sigma = O$ and thus a natural candidate would be to consider a representation $G_{K/k} \to \mathrm{Aut}(E[m])$. In the case $\mathrm{char}(k) \nmid m$ we can pick a basis for $E[m] = \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$ and identify $\mathrm{Aut}(E[m])$ with $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ obtaining thus a representation $G_{K/k} \to \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$. If we repeat this construction with $m = \ell^e$ for all values $e \in \mathbb{N}$ with ℓ prime, then the

intuition is to mimic the construction of the p-adics by showing that these homomorphisms are compatible under the $[l]$ map.

Definition 1.4.1. *Let ℓ be a prime and E an elliptic curve, the $\ell-$adic Tate module of E is $T_\ell(E) = \varprojlim_n E[\ell^n]$ where the inverse limit is taken with respect to the maps $E[\ell^{n+1}] \xrightarrow{\ell} E[\ell^n]$.*

In more detail if $P \in E : \ell^{n+1}P = 0$ then $\ell P \in E[\ell^n]$ and thus multiplication by ℓ is the same as reducing each point to it's representative in $E[\ell^n]$. Since each of the $E[\ell^n]$ is a $\mathbb{Z}/\ell^n\mathbb{Z}$ module the inverse limit inherits the structure of a $\mathbb{Z}_\ell$ module. This leads us to:

Proposition 1.4.2. *It holds:*

(i) $T_\ell(E) \simeq \mathbb{Z}_\ell \times \mathbb{Z}_\ell$ *if* $\ell \nmid \mathrm{char}(k)$

(ii) $T_\ell(E) = \mathbb{Z}_\ell$ *or* $\{0\}$ *otherwise*

Proof. The proof follows in both cases from Prop. 1.3.4. Pick generators P_n, Q_n for each $E[\ell^n]$ such that $\ell P_{i+1} = P_i$ and similarly for Q_i (this can always be done by picking a preimage of $E[\ell^{n+1}] \xrightarrow{\ell} E[\ell^n]$), then the elements of $T_\ell(E)$ are of the form $P = (a_1P_1, ..., a_nP_n, ...), Q = (b_1Q_1, ..., b_nQ_n, ...)$ with $a_i, b_i \in \mathbb{Z}/\ell^i\mathbb{Z}$ and $a_{i+1}\ell P_{i+1} = a_iP_i \implies a_{i+1}P_i = a_iP_i \iff a_{i+1} = a_i \mod \ell^i$ and similarly $b_{i+1} = b_i \mod \ell^i$ proving that $a = (a_1, ..., a_n, ..)$ and $b = (b_1, ..., b_n, ..)$ are in $\mathbb{Z}_\ell$. The required isomorphism is the one sending $(P, Q) \to (a, b)$. Case 2 is similar. □

The Galois group $G_{K/k}$ acts on T_ℓ as it commutes with the maps $E[\ell^{n+1}] \xrightarrow{\ell} E[\ell^n]$. With this settled we can now define the $\ell-$adic representation of $G_{K/k}$ associated to E.

Definition 1.4.3. *The $\ell-$adic representation is the homomorphism $\rho_{E,\ell} : G_{K/k} \to \mathrm{Aut}(T_\ell(E)) \cong \mathrm{GL}_2(\mathbb{Z}_\ell) \subset \mathrm{GL}_2(\mathbb{Q}_\ell)$.*

1.5 Weil Pairing

A key ingredient in order to further examine these representations is the Weil pairing. It is a bilinear, alternating pairing between points of $E[m]$. The easiest pairing with this property is simply to pick two basis points P, Q and for any two points $S = aP + bQ, T = cP + dQ$ to consider the discriminant $det(S, T) = ad - bc$. This however is not Galois invariant in the sense that $det(S, T)^\sigma \neq det(S^\sigma, T^\sigma)$ in general. The situation can be turned around however if instead we consider a pairing of the form $\zeta^{det(S,T)}$ with ζ a primitive $m-$th root of unity. The constuction is rather technical.
Let $T \in E$ and take a function $f \in k(E) : div(f) = mT - mO$ as well as the preimage of T under $[m]$, that is $T' \in E : [m]T' = T$. In a similar manner there is $g \in k(E) : div(g) = \sum_{R \in E[m]}(T + R) - R$. Notice that $div(f \circ [m]) = div(g^m)$ so we can assume equality up to a constant. Then if $S \in E[m]$, for any $X \in E$ we have $g(X + S)^m = f([m]X + [m]S) = f([m]X) = g(X)^m \iff (\frac{g(X+S)}{g(X)})^m = 1$.

Definition 1.5.1. *The Weil pairing (with the above notation) is defined as $e_m(S, T) = \frac{g(S+X)}{g(X)}, X \in E : g(X) \neq 0$ and it is a pairing $e_m : E[m] \times E[m] \to \mathbb{C}$.*

From the construction above it is immediate that e_m is a root of unity. The following proposition establishes the pairing's main properties.

Proposition 1.5.2. *The Weil pairing has the following properties:*

(i) *bilinearity:* $e_m(S_1 + S_2, T) = e_m(S_1, T)e_m(S_2, T)$ *and similarly for* T,

(ii) *it is alternating:* $e_m(S, T) = e_m(T, S)^{-1}$,

(iii) *Galois invariance:* $e_m(S, T)^\sigma = e_m(S^\sigma, T^\sigma)$,

(iv) *Compatibility:* $e_m m'(S, T) = e_m([m']S, T)$.

We will now extend this to a pairing of Tate modules. The only thing we need to show is compatibility with the maps $E[\ell^{n+1}] \xrightarrow{\ell} E[\ell^n]$. We thus need to show that $e_{\ell^{n+1}}(S, T)^\ell = e_{\ell^n}([\ell]S, [\ell]T)$. But linearity already gives us that $e_{\ell^{n+1}}(S, T)^\ell = e_{\ell^n+1}(S, [\ell]T) \overset{1.5.2iv}{=} e_{\ell^n}([\ell]S, [\ell]T)$. We have thus established that:

Theorem 1.5.3. *There exists an alternating, bilinear, non degenerate, Galois invariant pairing* $e : T_\ell(E) \times T_\ell(E) \to T_\ell(\mu)$ *where* μ *is an* ℓ *root of unity and the Tate module on the right hand side is* $T_\ell(\mu) = \varprojlim_n \mu_{\ell^n}$ *with respect to the maps* $\mu_{\ell^{n+1}} \overset{\zeta \to \zeta^\ell}{\to} \mu_{\ell^n}$.

1.6 Local fields and reduction properties

In this section we will examine elliptic curves over local fields and the different reduction types modulo a prime. We will first look at elliptic curves over finite fields. The most important morphism that comes into play here is the Frobenius endomorphism, that is $\sigma_p : E \to E$ with $\sigma_p(x, y) = (x^p, y^p)$. If we define our curve over a finite field $\mathbb{F}_p$ then $x^p = x, \forall x \in \mathbb{F}_p$ so the points in $\mathbb{F}_p$ are exactly those in the kernel of the map $\sigma_p - [1]$. This isogeny is easily seen to be seperable and thus $\#E(\mathbb{F}_p) = \deg(\sigma_p - [1])$. A famous result due to Hasse is the following inequality:

Theorem 1.6.1. ***(Hasse)*** *With the above notation:*

$$|\#E(\mathbb{F}_p) - p - 1| \leq 2\sqrt{p}$$

The quantity $a_p(E) = p + 1 - \#E(\mathbb{F}_p)$ is often referred to as the "trace of Frobenius" for reasons that will become apparent in the next section. Thinking of elliptic curves over finite fields is essentially the same as reducing the curve defined over a local field $\mod \pi$, where π is a uniformizer. Sometimes this approach does not provide us with an elliptic curve however as the resulting reduced curve $\tilde{E}$ can be singular. For example, the curve $y^2 = x^3 - p$ is easily seen to be singular $\mod p$ but nonsingular over $\mathbb{Q}$. Throughout the rest of this section let:

(i) K be a local field complete with respect to a valuation v,

(ii) R be the ring of integers of K,

(iii) R^* be the group of units,

(iv) π be a uniformizer for R,

(v) k is the residue field $R/\pi R$.

Suppose now that we are given a Weierstrass equation for the curve E/K. Since $i : (x, y) \to (u^{-2}x, u^{-3}y)$ leads to a new equation with u appearing in the coefficients we end up with a situation where we cannot talk about properties of the curve like if the discriminant in k is 0 or not unless we pick a specific equation. Since our isomorphism i scales the coefficients of our Weierstrass equation by a factor of u, we can assume $a_i \in R$ and $v(\Delta) \geq 0$. We thus use a Weierstrass model with minimal discriminant with respect to the constraint $a_i \in R$ and call it the "Minimal Weierstrass equation". If we pick a discrete valuation then $v(\Delta)$ is discrete and thus there exists a minimal equation for E/K. This equation is unique up to a change of coordinates like i with $u \in R^*$.

Let $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ be a minimal Weierstrass equation of E/K, then we define the reduction of E over k denoted $\tilde{E}$ as the curve defined by $y^2 + \tilde{a_1}xy + \tilde{a_3}y = x^3 + \tilde{a_2}x^2 + \tilde{a_4}x + \tilde{a_6}$ over the residue field k. This defines a reduction map $E(K) \to \tilde{E}(k)$ which can be shown to be surjective using Hensel's Lemma. Let $\tilde{E}_{ns}(k)$ denote the non singular points of the reduced curve $\tilde{E}(k)$. We define two sets $E_0(K) = \{P \in E(K) : \tilde{P} \in \tilde{E}_{ns}(k)\}$ and $E_1(K) = \{P \in E(K) : \tilde{P} = O\}$. We can prove that $E_0(K)$ is actually a group and then deduce that:

Proposition 1.6.2. *There is an exact sequence:*
$0 \to E_1(K) \to E_0(K) \to \tilde{E}_{ns} \to 0$.

There is also a key property in the reduction of the torsion subgroup $E[m]$.

Proposition 1.6.3. *Let E/K be an elliptic curve and m relatively prime to* $\mathrm{char}(k)$. *Then:*

(i) *$E_1(K)$ contains no non-trivial points of order m,*

(ii) *if the reduced curve is non singular then the reduction $E(K)[m] \to \tilde{E}(k)$ is injective.*

This is a weak version of one direction of the criterion of Néron–Ogg–Shafarevich which we will see later.
We will now shift our attention to the possible reduction types of an elliptic curve.

Definition 1.6.4. *Let E/K be an elliptic curve and $\tilde{E}$ it's reduction over $k = R/(\pi)$. We say that E has:*

(i) *good reduction* $\mod \pi$ *if $\tilde{E}$ is non singular* $\iff v(\Delta) = 0$

(ii) *multiplicative reduction if $\tilde{E}(k)$ has a nodal singularity* $\iff v(\Delta) > 0$ *and* $v(c_4) = 0$

(iii) *additive reduction if $\tilde{E}(k)$ has a cuspidal singularity* $\iff v(\Delta) > 0$ *and* $v(c_4) > 0$

There is also another type of reduction we will be referring to and that is potentially good reduction.

Definition 1.6.5. *We say E/K has potentially good reduction if there exists an extension $F \supset K$ such that E/F has good reduction in the corresponding residue field.*

A property we will be using is the following:

Proposition 1.6.6. *E/K has potentially good reduction if and only if $j(E) \in R$.*

In the rational case, given all the local information it is possible to minimize the discriminant with respect to every valuation v_p simultaneously giving us the "global minimal discriminant" of E over $\mathbb{Q}$. So given the minimal discriminant of an elliptic curve $\Delta(E) = \prod_{i=1}^{n} p_i^{e_i}$ we can immediately tell the primes of good and bad reduction by looking at which primes divide the discriminant. We also define a really important quantity which will later be our way to identify the space of eigenforms of an elliptic curve.

Definition 1.6.7. *We define the conductor N of an elliptic curve defined over a number field K as the prime ideal that is divisible by exactly the prime ideals where E has bad reduction. We write $N = \prod_{i=1}^{n} \mathfrak{p}_i^{\mathfrak{e}_i}$ where $\mathfrak{e}_i = 1$ if E has multiplicative reduction over $\mathfrak{p}_i$, $\mathfrak{e}_i = 2$ if E has additive reduction and $\mathfrak{p}_i \nmid 2$ or 3. The cases $2, 3$ are treated seperately and $\mathfrak{e}_i$ is given by Tate's algorithm (p.364 [3.3]).*

Tate's Algorithm will be used to compute the minimal discriminant and conductor of our specific Frey curves in Chapter 3. It takes as input an elliptic curve over a local field with a uniformizer π (in our case $\mathbb{Q}_p$ and p) in the form of a Weierstrass equation and outputs $v_p(\Delta_{min}), v_p(N)$ where D_{min}, N are the global minimal discriminant and the conductor. The conductor thus encodes all the places of bad reduction of our curve and the reduction type specifically.
We will now conclude this section with a result more about the Frobenius automorphism in the case of good reduction.

Proposition 1.6.8. *Let $E/\mathbb{Q}$ be an elliptic curve with good reduction at p. If $\widehat{\sigma_p}$ is the dual of the Forbenius endomorphism σ_p then $[a_p(E)] = \sigma_p + \widehat{\sigma_p}$.*

Proof. We already established that $\#E(\mathbb{F}_p) = \deg(1 - \sigma_p) = (1 - \sigma_p)(\widehat{1 - \sigma_p}) = 1 - \sigma_p - \widehat{\sigma_p} + \sigma_p \circ \widehat{\sigma_p} = 1 + \deg(\sigma_p) - \sigma_p - \widehat{\sigma_p} = 1 + p - \sigma_p - \widehat{\sigma_p} \implies a_p(E) = \sigma_p + \widehat{\sigma_p}$ □

1.7 L-functions of elliptic curves

In the previous section we examined elliptic curves over finite fields and presented Hasse's result on bounds of $a_p(E)$. Attaching an L-function to an elliptic curve is mainly about obtaining information about the curve as a variety defined over progressively larger fields. More specifically, the idea is to attach a series that depends on the number of solutions in $\mathbb{F}_{p^n}$.

Definition 1.7.1. *Let E be an elliptic curve over $\mathbb{Q}$. The zeta function of E over $\mathbb{F}_p$ is defined to be the power series:*

$$Z(E/\mathbb{F}_p, T) = exp(\sum_{n=0}^{\infty} (\#E(\mathbb{F}_{p^n})) \frac{T^n}{n})$$

Notice that from this definition we can obtain the number of elements over each finite field extension by:

$$\#E(\mathbb{F}_{p^n})) = \frac{1}{(n-1)!}\frac{d^n}{dT^n} log(Z(E/\mathbb{F}_p, T))|_{T=0}$$

A famous conjecture by Weil asserts several facts that should be true for such functions for a general projective variety. In the case of elliptic curves we have that:

Theorem 1.7.2. *If $E/\mathbb{F}_p$ is an elliptic curve, then it holds:*

$$Z(E/\mathbb{F}_p, T) = \frac{L_p(T)}{(1-T)(1-pT)},$$

where $L_p(T) = 1 - a_p(E)T + pT^2$

We expand the definition of this term L_p in the case of E not having good reduction at p:

$$L_p = \begin{cases} 1 - a_p(E)T + pT^2, & \text{E has good reduction at } p \\ 1 - T, & \text{E has split multiplicative reduction at } p \\ 1 + T, & \text{E has non-split multiplicative reduction at } p \\ 1, & \text{E has additive reduction at } p. \end{cases}$$

Definition 1.7.3. *The L-series of an elliptic curve $E/\mathbb{F}_p$ is defined as:*

$$L(E, s) = \prod_p \frac{1}{L_p(p^{-s})},$$

for all primes p.

From Hasse's inequality it follows that this product converges and defines an analytic function for all $Re(s) > \frac{3}{2}$. We will see in the next chapter a result linking this function to the $L-$function of a modular form which will immediately imply the following result:

Theorem 1.7.4. ***(Hasse-Weil Conjecture)*** *Let $E/\mathbb{Q}$ be an elliptic curve, then the function $L(E, s)$ satisfies a functional equation linking $\Lambda(E, s)$ and $\Lambda(E, 2-s)$. The same theorem also allows us to analytically continue $L(E, s)$ on all of $\mathbb{C}$.*

1.8 Elliptic Curve Representations

Equipped with the Weil pairing we can now actually calculate the action of the Galois group explicitly if we know how it acts on a basis as the next proposition suggests. We set $G_{\mathbb{Q}} = Gal(\bar{\mathbb{Q}}/\mathbb{Q})$, the absolute Galois group of the rationals.

Proposition 1.8.1. *Let $\rho_{E,\ell} : G_{\mathbb{Q}} \to \mathrm{Aut}(T_\ell(E))$ be a representation as before and denote with $\rho_n : G_{\mathbb{Q}} \to \mathrm{Aut}(\mathbb{Z}/\ell^n\mathbb{Z})$ it's $n-th$ entry. Then for every $\sigma \in G$ we have $\mu_{\ell^n}^{\sigma} = \mu_{\ell^n}^{det(\rho_n(\sigma))}, \forall n \geq 1$.*

Proof. Let P_n, Q_n be a basis for $\mathrm{Aut}(\mathbb{Z}/\ell^n\mathbb{Z})$. Then $[P_n^\sigma, Q_n^\sigma] = \rho_n(\sigma)\begin{bmatrix}P\\Q\end{bmatrix} = [aP_n + bQ_n, cP_n + dQ_n]$ for some $a, b, c, d \in \mathbb{Z}/\ell^n\mathbb{Z}$ From Galois invariance we have that $e_{\ell^n}(P_n^\sigma, Q_n^\sigma) = e_{\ell^n}(P_n, Q_n)^\sigma = \mu_{\ell^n}^\sigma$. But $e_{\ell^n}(P_n^\sigma, Q_n^\sigma) = e_{\ell^n}(aP_n + bQ_n, cP_n + dQ_n) = e_{\ell^n}(aP_n, cP_n)e_{\ell^n}(aP_n, dQ_n)e_{\ell^n}(bQ_n, cP_n)e_{\ell^n}(bQ_n, dQ_n) = e_{\ell^n}(P_n, Q_n)^{det(\rho_n(\sigma))} = \mu_{\ell^n}^{det(\rho_n(\sigma))}$. □

We now look a bit deeper at the structure of $G_\mathbb{Q}$. For every number field extension $F/\mathbb{Q}$ we have a surjection $G_\mathbb{Q} \to Gal(F/\mathbb{Q})$ called the restriction of an automorphism at F and denoted $\sigma_F, \sigma \in G_\mathbb{Q}$. Let $p \in \mathbb{Z}$ be a prime and $\mathfrak{p}$ a maximal ideal over it in $\bar{\mathbb{Z}}$. We then have the following short exact sequence

$$0 \to I_\mathfrak{p} \to D_\mathfrak{p} \to Gal((\bar{\mathbb{Z}}/(\mathfrak{p}))/(\mathbb{Z}/(p))) \to 0$$

where $D_\mathfrak{p} = \{\sigma \in G_\mathbb{Q} : \mathfrak{p}^\sigma = \mathfrak{p}\}$ the decomposition group and $I_\mathfrak{p} = \{\sigma \in G_\mathbb{Q} : a^\sigma = a \mod \mathfrak{p}\}$ is the inertia group of the extension. In this way we obtain an isomorphism $D_\mathfrak{p} \to Gal(\overline{\mathbb{F}}_p/\mathbb{F}_p)$.
The next object we will study is the Frobenius elements of our extension as they will turn out to be of significant importance in the study of the associated represenations. In the context of a finite extension we have similarly: $i : D_\mathfrak{p}/I_\mathfrak{p} \cong Gal(F_p/f_p)$ where F_p, f_p are the corresponding residue fields. We then have that $Gal(F_p/f_p) = \langle\sigma_p\rangle$ where $\sigma_p(x) = x^p$ is the Frobenius automorphism. A Frobenius element is then any representative of this element in $D_\mathfrak{p}$. In more detail:

Definition 1.8.2. *Let $F/\mathbb{Q}$ be a Galois extension, $p \in \mathbb{Z}$ a prime and $\mathfrak{p}$ a corresponding maximal ideal in $\mathcal{O}_F$. The Frobenius element of $Gal(F/\mathbb{Q})$ is an element $Frob_\mathfrak{p}$ satisfying the condition $x^{Frob_\mathfrak{p}} = x^p \mod \mathfrak{p}$. Up to conjugation the Frobenius element is independant of $\mathfrak{p}$ and can be written $Frob_p$.*

Similarly in the case of $G_\mathbb{Q}$ the absolute Frobenius element is a preimage of the Frobenius automorphism $\sigma_p \in Gal(\overline{\mathbb{F}_p}/\mathbb{F}_p)$. Observe that the absolute Frobenius element is again specified up to the inertia group as $\sigma I_\mathfrak{p}$. We are only interested in $\rho_{G_\mathbb{Q}}(Frob_\mathfrak{p})$ however so in order for this to make sense we naturally need $I_\mathfrak{p} \subseteq ker(\rho_{G_\mathbb{Q}})$. This motivates the following:

Definition 1.8.3. *Let ρ be a Galois representation and p a prime. Then we say ρ is unramified at p if $I_\mathfrak{p} \subset ker(\rho)$ for every maximal ideal $p \subset \mathfrak{p}$.*

Notice that we defined $\ell-$adic representations but now we talk about representations $G_\mathbb{Q}$. This is were our next definition emerges from as it is a way to write down automorphisms in the $\ell-$adic setting of $T_\ell(\mu)$.

Definition 1.8.4. *Let μ_{ℓ^n} be an ℓ^n-th root of unity. Then we define the $\ell-$adic cyclotomic character as the one dimensional represenation $\chi_\ell : G_\mathbb{Q} \to \mathbb{Q}_\ell^*$ defined by $\sigma \to (m_1, ..., m_n) : \mu_{\ell^n}^\sigma = \mu_{\ell^n}^{m_n}, \forall n$.*

As an immdiate result one gets the following:

Proposition 1.8.5. *With the previous notation if $p \nmid \ell$ then $\chi_\ell(Frob_\mathfrak{p}) = p$.*

Proof. From the definition of a Frobenius element we have that $\mu_{\ell^n}^{Frob_\mathfrak{p}} = \mu_{\ell^n}^p \mod \mathfrak{p} \implies \mathfrak{p} | \mu_{\ell^n}^p(\mu_{\ell^n}^{Frob_\mathfrak{p} - p} - 1)$ but if $Frob_\mathfrak{p} - p \neq 0 \mod \ell^n$ then taking norms

we have that $N(\mu_{\ell^n}^{Frob_\mathfrak{p}-p}-1)=\ell$ (by the cyclotomic polynomial $(X+1)^{\ell^n}=1$) and $N(\mu_{\ell^n}^{p})=1$ implying that $N(\mathfrak{p})=p|\ell$ which is a contradiction. We thus obtain that $Frob_\mathfrak{p}-p=0 \mod \ell^n, \forall n \implies \chi_\ell(Frob_\mathfrak{p})=(p,p,p,...)=p$. □

We will finally use a well known result that connects the reduction properties of an elliptic curve with the ramification properties of the associated Galois representation. We call a module unramified at the uniformizer π if the action of the inertia group I_π of $G_{\bar{K}/K}$ on it is trivial.

Theorem 1.8.6. *(**Criterion of Néron–Ogg–Shafarevich**) Let E/K be an elliptic curve then the following statements are equivalent:*

(i) *E has good reduction at K,*

(ii) *$E[m]$ is unramified at π for every m coprime to* $\mathrm{char}(k)$,

(iii) *$T_\ell(E)$ is unramified at π for every $\ell \neq \mathrm{char}(k)$.*

The above combine with the preview section's results to give the following description of the characteristic polynomial of a Frobenius element. When we discuss modular curves and their attached representations this theorem will allow us to transition from the representation similarity version of the Modularity Theorem to that of $q-$ expansions.

Theorem 1.8.7. *Let ℓ be a prime and E an elliptic curve with conductor N. The Galois representation $\rho_{E,\ell}$ is then unramified at all primes $p \nmid \ell N$ and for any such p if $\mathfrak{p}$ is a maximal ideal of $\mathbb{Z}$ lying over p, then the characteristic polynomial of $\rho_{E,\ell}(Frob_\mathfrak{p})$ is $x^2-a_p(E)x+p$. Moreover the Galois representation is irreducible.*

Proof. The extension is unramified by ii) of the above criterion if and only if it has good reduction at p, that is $p \nmid N$ and $\ell \neq \mathrm{char}(k)=p \iff p \nmid \ell N$. We now observe that the characteristic polynomial is simply $x^2-tr(\rho_{E,\ell}(Frob_\mathfrak{p}))x+det(\rho_{E,\ell}(Frob_\mathfrak{p}))$. Let $\rho_n : G_\mathbb{Q} \to \mathrm{GL}_2(\mathbb{Z}/\ell^n\mathbb{Z})$ be the restriction of the action of $\rho_{E,\ell}$ as before. We have already established by Proposition 1.8.1 that $\mu_{\ell^n}^{\sigma}=\mu_{\ell^n}^{det(\sigma)}=\mu_{\ell^n}^{\chi_{\ell,n}(\sigma)}$ as by definition the action is to raise μ_{ℓ^n} to the $n-$th term of it's cyclotomic character $\chi_\ell(\sigma)$. We thus obtain $det(\rho_n(\sigma))=\chi_{\ell,n}$ for every n which implies $det(\rho_{E,\ell}(Frob_\mathfrak{p}))=\chi_\ell(Frob_\mathfrak{p})=p$. Let $A=\rho_{E,\ell}(Frob_\mathfrak{p})$. Substituing in the characteristic polynomial (every 2×2 matrix satisfies it's characteristic equation) we get $tr(A)I=A+pA^{-1}$ and thus it is enough to show that $a_p(E)I=A+pA^{-1}$. But $\rho_{E,\ell}(\sigma_p\widehat{\sigma_p})=\rho_{E,\ell}(p)=pI$ as we are simply multiplying by p. Then $\rho_{E,\ell}(\sigma_p)\rho_{E,\ell}(\widehat{\sigma_p})=pI \implies \rho_{E,\ell}(\widehat{\sigma_p})=p\rho_{E,\ell}(\sigma_p)^{-1}$. Now from Proposition 1.6.8 we get $[a_p(E)]=\sigma_p+\widehat{\sigma_p}$ and applying $\rho_{E,\ell}$ to both sides we get $a_p(E)I=A+pA^{-1}=tr(A)I$ as desired. The proof of irreducibility is out of the scope of this thesis. □

Chapter 2

Modular Forms

This chapter is devoted to the study of modular forms as they ply a key role in the modular method. Modular forms are essentially functions on the complex plane that exhibit several remarkable properties of symmetry, so many in fact that many mathematicians consider it almost a "welcome coincidence" that they even exist! We will naturally begin with the basic definitions.

2.1 Basic Definitions

The first question that probably comes to mind when we consistently talk about the outstanding symmetry properties of modular forms is "what are they symmetric with respect to?". There are different classes of modular forms that can be defined using different subgroups of $\mathrm{SL}_2(\mathbb{Z})$.

Definition 2.1.1. *Let N be a positive integer. We define the principal congruence subgroup of level N as:* $\Gamma(N) = \{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \mod N\}$. *We call a subgroup of Γ a congruence subgroup of level N if $\exists N \in \mathbb{N}$ such that $\Gamma(N) \subseteq \Gamma$.*

There are two important cases of congruence subgroups:

(i) $\Gamma_0(N) = \{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} * & * \\ 0 & * \end{bmatrix} \mod N\}$

(ii) $\Gamma_1(N) = \{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & * \\ 0 & 1 \end{bmatrix} \mod N\}$

where the $*$ means that we can have any number in that place.

Remark 2.1.2. *Notice that any congruence subgroup Γ contains some $\Gamma(N)$ and thus $|\mathrm{SL}_2(\mathbb{Z}) : \Gamma| \leq |\mathrm{SL}_2(\mathbb{Z}) : \Gamma(N)|$ which is finite. The same holds for the index of any two congruence subgroups.*

Definition 2.1.3. *Let $\mathbb{H} = \{\tau \in \mathbb{C} : Im(\tau) > 0\}$ be the upper half plane. For $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ we define the factor of automorphy $j(\gamma, \tau) = c\tau + d$ and the weight k operator $[\gamma]_k$ on a function $f : \mathbb{H} \to \mathbb{C}$ to be $(f[\gamma]_k)\tau) = j(\gamma, \tau)^{-k} f((\gamma(\tau)))$ and $\gamma(\tau) = \frac{a\tau+b}{c\tau+d}$.*

We can now finally define what a modular form is:

Definition 2.1.4. *Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. A function $f : \mathbb{H} \to \mathbb{C}$ is a modular form of weight k with respect to Γ if:*

(i) *f is holomorphic*

(ii) *$\forall \gamma \in \Gamma$ we have $f[\gamma]_k = f$*

(iii) *$f[a]_k$ is holomorphic at $\infty, \forall a \in \mathrm{SL}_2(\mathbb{Z})$*

We denote the vector space over $\mathbb{C}$ of modular forms of weight k with respect to Γ as $\mathcal{M}_k(\Gamma)$. We also define $\mathcal{M}(\Gamma) = \bigoplus_{k\in\mathbb{Z}} \mathcal{M}_k(\Gamma)$ the graded ring formed by these vector spaces of functions. Notice that since every congruence subgroup contains $\Gamma(N)$, then it must contain a translation element of the form $h = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix}$. This means that if f is a modular form with respect to Γ, then it has a Fourier expansion $f(\tau) = \sum_{n=0}^{\infty} a_n q_h^n$ where $q_h = e^{2\pi i \tau/h}$. Thus f can be viewed as a holomorphic function with the punctured unit disc $D\backslash\{0\}$ (where $D = \{z \in \mathbb{C} : |z| < 1\}$) as it's domain. The third condition now implies that f can be extended to a holomorphic function on all of D.
We now define a subspace of modular forms that contains the objects we will actually attach to elliptic curves called newforms. This space is the space of cuspforms.

Definition 2.1.5. *A cusp form of weight k with respect to Γ is a modular form with the same weight and congruent subgroup such that the Fourier expansion of $f[a]_k$ has $a_0 = 0, \forall a \in \mathrm{SL}_2(\mathbb{Z})$. We similarly denote the set of cuspforms of weight k defined over Γ as $\mathcal{S}_k(\Gamma)$ and the graded space $\mathcal{S}(\Gamma) = \bigoplus_{k\in\mathbb{Z}} \mathcal{S}_k(\Gamma)$*

2.2 Modular curves and Dimension

Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Then the quotient space obtain via the action of Γ on $\mathbb{H}$ denoted $\Gamma\backslash\mathbb{H}$ is the modular curve $Y(\Gamma) = \{\Gamma\tau, \tau \in \mathbb{H}\}$. This space can be shown to be Hausdorff. This space is indeed a Riemann surface. There are certain points that need to be addressed specifically on this space called elliptic points.

Definition 2.2.1. *An elliptic point is a point $\tau \in Y(\Gamma)$ such that the isotropy group of τ denoted by $\Gamma_\tau = \{\gamma \in \Gamma : \gamma(\tau) = \tau\}$ is non trivial, that is $\Gamma_\tau \neq \{I\}$. If $\pi : \mathbb{H} \to Y(\Gamma)$ is the projection map, then $\pi(\tau)$ is also called elliptic.*

Thankfully the set of elliptic points is discrete as can be seen in Corollary 2.2.3 [3.3]. Also Γ_τ is finite cyclic by Corollary 2.2.5 [3.3]. These points determine (along with the cusps which we will see) below the dimenion of the space of modular forms. We thus try to identify them: Let $\gamma\tau = \tau$, then in matrix form we get $a\tau + b = c\tau^2 + d\tau \implies c\tau^2 + (a+d)\tau + b = 0$ and $|a+d| < 2$ since $Im(\tau) > 0$. The characteristic polynomial of γ is then $x^2 + 1$ or $x^2 + x + 1$ or $x^2 - x + 1$ which implies that γ has order $1, 2, 3, 4$ or 6 where only $3, 4$ and 6 correspond to $\gamma \neq \pm I$.

We now look at the second set of points that allow us to compute the dimension of a modular curve and these are the cusps. These points are used in order to compactify $Y(\Gamma)$ to a Riemann surface $X(\Gamma)$ which we can study with the known theory. We are specifically interested in the genus of $X(\Gamma)$.

Definition 2.2.2. *The cusps of* Γ *are the* Γ *equivalence classes of* $\mathbb{Q} \cup \{\infty\}$.

Since $s \in \mathbb{Q}$ takes the form $s = a(\infty), a \in \mathrm{SL}_2(\mathbb{Z})$, an upper bound for the number of cusps is $[\mathrm{SL}_2(\mathbb{Z}) : \Gamma]$ which is finite. To compute the compactification of the modular curve we add these points to our curve to get $X(\Gamma) = Y(\Gamma) \cup \Gamma\backslash(\mathbb{Q} \cup \{\infty\})$. In a similar way we can define the curves $X_0(\Gamma), X_1(\Gamma)$ corresponding to $\Gamma_0(N)$ and $\Gamma_1(N)$.

We can now specify the genus of $X(\Gamma)$ for a level N congruence subgroup.The standard proof uses the Riemann-Hurwitz formula.

Theorem 2.2.3. ***(Riemann-Hurwitz)*** *Let* $f : X \to Y$ *be a non-constant holomorphic map of Riemann surfaces and suppose* e_x *denotes the ramification degree of* f *at the point* x*. Then we have:*

$$2g_X - 2 = d(2g_Y - 2) + \sum_{x \in X} (e^x - 1),$$

where g_X *is the genus of* X *and similarly for* g_Y*.*

Let $\Gamma_1 \subseteq \Gamma_2$ be congruence subgroups and consider the natural projection map $f : X(\Gamma_1) \to X(\Gamma_2)$ given by $\Gamma_1\tau \to \Gamma_2\tau$. Consider now the corresponding isotropic groups $\Gamma_{1,\tau}, \Gamma_{2,\tau}$ and define $h_i = |\{\pm I\}\Gamma_{i,\tau}|/2$ to be the periods of the elliptic point τ. Then $h_i = 1, 2$ or 3 and we get the following result:

Theorem 2.2.4. *Let* $f : X(\Gamma) \to X(1)$ *be the natural projection with degree* d *and* e_2, e_3 *be the number of elliptic points with period* 2 *and* 3 *respectively and* e_∞ *the number of cusps. Then the genus of* $X(\Gamma)$ *is*

$$g = 1 + \frac{d}{12} - \frac{e_2}{4} - \frac{e_3}{3} - \frac{e_\infty}{2}.$$

This finally leads us to the complete characterization of the dimension of the space $\mathcal{M}_k(\Gamma)$.

Theorem 2.2.5. *For* k *an even integer, with notation as in the previous theorem we have that:*

$$dim_{\mathbb{C}}(\mathcal{M}_k(\Gamma)) = \begin{cases} (k-1)(g-1) + \lfloor \frac{k}{4} \rfloor e_2 + \lfloor \frac{k}{3} \rfloor e_3 + \frac{k}{2} e_\infty, & if k \geq 2 \\ 1 \quad , & if k = 0 \\ 0 \quad , & if k < 0 \end{cases}$$

and

$$dim_{\mathbb{C}}(\mathcal{S}_k(\Gamma)) = \begin{cases} (k-1)(g-1) + \lfloor \frac{k}{4} \rfloor e_2 + \lfloor \frac{k}{3} \rfloor e_3 + (\frac{k}{2} - 1) e_\infty, & if k \geq 4 \\ 1, & if k = 2 \\ 0, & if k \leq 0. \end{cases}$$

In particular the space of modular forms of weight k has finite dimension. In the odd case we can observe that if $-I \in \Gamma$ then $\mathcal{M}_k(\Gamma) = \{0\}$.

2.3 Hecke Operators

The first instances of Hecke operators were used by Mordell in his resolution of the Ramanujan conjecture and the exact objects were later studied by Hecke. The importance of Hecke operators lies in the fact that $\mathcal{S}_k(\Gamma)$ has a basis comprised of simultaneous eigenfunctions for these operators.

2.3a Double coset operators

The standard way to define Hecke operators is with the use of double coset operators which are characterized by the action of $GL_2^+(\mathbb{Q})$ on modular forms. The action on $\mathbb{H}$ is the same as that of $\mathrm{SL}_2(\mathbb{Q})$ and the same holds true for the factor of automorphy j. This time however we also need to acount for the discriminant so we define the $[\gamma]_k$ operator as:

$$(f([\gamma]_k))(\tau) = (det\gamma)^{k-1} j(\gamma,\tau)^{-k} f(\gamma(\tau))$$

It is straightforward to show the following.

Lemma 2.3.1. *It holds:*

(i) $j(\gamma\gamma',\tau) = j(\gamma,\gamma'(\tau))j(\gamma',\tau)$,

(ii) $\gamma\gamma'(\tau) = \gamma((\gamma'(\tau)))$,

(iii) $[\gamma\gamma']_k = [\gamma]_k[\gamma']_k$.

With this in mind we can now define a double coset of two congruence subgroups.

Definition 2.3.2. *For congruence subgroups $\Gamma_{1,2}$ of $\mathrm{SL}_2(\mathbb{Z})$ and $a \in GL_2^+(\mathbb{Q})$ we define the double coset as $\Gamma_1 a \Gamma_2 = \{\gamma_1 a \gamma_2, \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2\}$.*

Observe that the group Γ_1 acts by left multiplication on $\Gamma_1 a \Gamma_2$ thus partitioning it into orbits as $\Gamma_1 a \Gamma_2 = \cup \Gamma_1 \beta_j$ with representatives $\beta_j = \gamma_1 a \gamma_2$ for each orbit. We will show below that in fact this is a finite partition using the following two lemmas:

Lemma 2.3.3. *If Γ is a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and $a \in GL_2^+(\mathbb{Q})$ then $a^{-1}\Gamma a \cap \mathrm{SL}_2(\mathbb{Z})$ is also a congruence subroup of $\mathrm{SL}_2(\mathbb{Z})$.*

Lemma 2.3.4. *Let Γ_1 and Γ_2 be congrence subgroups and $a \in GL_2^+(\mathbb{Q})$. Set $\Gamma_3 = a^{-1}\Gamma_1 a \cap \Gamma_2$. Then there is a natural bijection from $\Gamma_3 \backslash \Gamma_2$ to the orbit space $\Gamma_1 \backslash \Gamma_1 a \Gamma_2$ defined by $\gamma_2 \to a\gamma_2$ which takes $\Gamma_2 \to \Gamma_1 a \Gamma_2$.*

Proof. Immediate since for $\gamma_2, \gamma_2' \in \Gamma_2$ are taken to same orbit $\Gamma_1 a \gamma_2 = \Gamma_1 a \gamma_2 \iff \gamma_2(\gamma_2')^{-1} \in a^{-1}\Gamma_1 a$. □

From the above lemma it is immediately observed that it suffices to show that $|\Gamma_3 : \Gamma_2|$ is finite which always holds as long as we can show that Γ_3 is a congruence subgroup which is established by Lemma 2.3.3 and the fact that if $\Gamma(n) \subseteq \Gamma_1$ and $\Gamma(m) \subseteq \Gamma_2$ then $\Gamma(lcm(m,n)) \subseteq \Gamma_1 \cap \Gamma_2$ making $\Gamma_1 \cap \Gamma_2$ a congruence subgroup. We thus proved that:

Proposition 2.3.5. *The action of Γ_1 on $\Gamma_1 a \Gamma_2$ partitions it into finitely many orbits $\Gamma_1 a \Gamma_2 = \cup_{j=1}^n \Gamma_1 \beta_j$.*

We wil now define the double coset operator.

Definition 2.3.6. *Let Γ_1 and Γ_2 be congruence subgroups and $a \in GL_2^+(\mathbb{Q})$ with $\Gamma_1 a \Gamma_2 = \cup_{j=1}^n \Gamma_1 \beta_j$. We define the k-weight double coset operator as $[\Gamma_1 a \Gamma_2]_k : \mathcal{M}_k(\Gamma_1) \to \mathcal{M}_k(\Gamma_2)$ with $f[\Gamma_1 a \Gamma_2]_k = \sum_{j=1}^n f[\beta_j]_k$.*

Observe that the reason this is indeed a map $\mathcal{M}_k(\Gamma_1) \to \mathcal{M}_k(\Gamma_2)$ is because if β_j are the orbit representatives and $\gamma_2 \in \Gamma_2$ then $\beta_j\gamma_2$ are also orbit representatives and thus $(f[\Gamma_1 a\Gamma_2]_k)([\gamma_2]_k) = f[\Gamma_1 a\Gamma_2]_k$. Holomorphy is also immediate as it is the finite sum of holomorphic functions. Specifically if these functions vanish at infinity then so does $f[\Gamma_1 a\Gamma_2]_k$ giving us a map $\mathcal{S}_k(\Gamma_1) \to \mathcal{S}_k(\Gamma_2)$.

2.3b Diamond and T_p operators

We now focus on modular forms on $\Gamma_1(N)$ and give two types of Hecke operators that are of central importance to us.

Definition 2.3.7. *For $d \in \mathbb{Z}$ and $\alpha \in \Gamma_0(N)$ with $\alpha = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}$ such that $d = a_{2,2} \mod N$ define the Diamond operator $\langle d\rangle : \mathcal{M}_k(\Gamma_1(N)) \to \mathcal{M}_k(\Gamma_1(N))$ as $\langle d\rangle f = f[\alpha]_k$.*

Notice that we have an isomorphism $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^*$ given by $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \to d \mod N$. This means that the operator $\langle d\rangle$ is well defined.

Definition 2.3.8. *The second type of Hecke operator is defined as $T_p : \mathcal{M}_k(\Gamma_1(N)) \to \mathcal{M}_k(\Gamma_1(N)$ with $T_p f = f[\Gamma_1(N) \begin{bmatrix} 1 & 0 \\ 0 & p \end{bmatrix} \Gamma_1(N)]_k$.*

Considering different divisibility cases we can find an explicit formula for this double coset operator by identifying the orbit representatives in each case.

Proposition 2.3.9. *For T_p as above, we have:*

(i) $T_p f = \sum_{j=1}^{p-1} f \begin{bmatrix} 1 & j \\ 0 & p \end{bmatrix}_k$ *if* $p|N$

(ii) $T_p f = \sum_{j=1}^{p-1} f \begin{bmatrix} 1 & j \\ 0 & p \end{bmatrix}_k + f \begin{bmatrix} m & n \\ N & p \end{bmatrix}_k \begin{bmatrix} 1 & 0 \\ 0 & p \end{bmatrix}_k$ *if* $p \nmid N$ *and* $mp - nN = 1$

There is more that can be said about the way a Hecke operator acts on the Fourier expansions of modular forms.

Definition 2.3.10. *Let $\chi : (\mathbb{Z}/N\mathbb{Z})^* \to \mathbb{C}^*$ be a character and denote $\mathcal{M}_k(N, \chi) = \{f \in \mathcal{M}_k(\Gamma_1(N)) : \langle d\rangle = \chi(d)f, \forall d \in \mathbb{Z}/N\mathbb{Z})^*\}$*

Proposition 2.3.11. *T_p preserves $\mathcal{M}_k(N, \chi)$ and if $f \in \mathcal{M}_k(N, \chi)$, it's Fourier expansion is $(T_p f)(\tau) = \sum_{n=0}^{\infty}(a_{np}(f) + \chi(p)p^{k-1}a_{n/p}(f))q^n$ with $a_{n/p} = 0$ if $p \nmid n$. In particularl for $f \in \mathcal{S}_k(N)$ we get that $T_p f \in \mathcal{S}_k(N)$.*

So what this proposition tells us is that applying T_p to f causes the Fourier coefficients to change as $a_n(T_p f) = a_{np}(f) + \chi(p)p^{k-1}a_{n/p}(f)$. Another really important property of these Hecke operators is that they commute. This will allow us to use spectral type theorems on the complex vector spaces of modular forms.

Proposition 2.3.12. *For the Diamond operator $\langle d\rangle$ and T_p we have that:*

(i) $\langle d\rangle T_p = T_p\langle d\rangle$,

(ii) $\langle d\rangle\langle e\rangle = \langle e\rangle\langle d\rangle$,

(iii) $T_pT_q = T_qT_p$.

Using the above, there is a way to define T_n for any $n \in \mathbb{N}$: we first define it for powers of primes: $T_{p^r} = T_pT_{p^{r-1}} - p^{k-1}\langle p\rangle T_{p^{r-2}}$ and $T_n = \prod_{j=1}^{n} T_{p_j^{r_j}}$ by commutativity, where we simply take the prime decomposition $n = \prod_{j=1}^{n} p_j^{r_j}$.

Definition 2.3.13. *The $\mathbb{C}-$algebra generated by the Hecke operators $\langle n\rangle$ and T_n is called the Hecke algebra over $\mathbb{C}$ and denoted $\mathbb{T}_{\mathbb{C}}$.*

2.3c Petersson Inner Product

In this section we will examine more in depth the space $\mathcal{S}_k(\Gamma_1(N))$ of cuspforms. We will equip this space with an inner product which will allow us to treat it as an inner product space with Hermitian operators and thus show that there is a basis that is also a simultaneous eigenvalue for all these Hecke operators we saw.

Let $\tau = x+yi$ lie in $\mathbb{H}$ and consider the measure $d\mu(\tau) = \frac{dxdy}{y^2}$. This measure is easily observed to be not just $\mathrm{SL}_2(\mathbb{Z})$ invariant but $GL_2^+(\mathbb{Q})$ invariant. A fundamental domain of $\mathbb{H}^*$ under the action of $\mathrm{SL}_2(\mathbb{Z})$ is $\mathbb{D} = \{\tau \in \mathbb{H} : \Re(\tau) \leq \frac{1}{2}, |\tau| \geq 1\}$. For a congruence subgroup Γ we write $\mathrm{SL}_2(\mathbb{Z}) = \cup_j\{\pm I\}\Gamma a_j$ as a finite set of cosets. Observe that $\Gamma\backslash\mathbb{H} = \Gamma\backslash(\cup_j\mathbb{D}a_j)$ and thus we get that up to boundary identification $X(\Gamma)$ can be represented as $\cup_j\mathbb{D}a_j$. We thus obtain by integrating and noticing that $\forall f \in \mathcal{S}_k(\Gamma), f$ is Γ invariant that $\int_{X(\Gamma)} f d\mu(\tau) = \int_{\cup_j\mathbb{D}a_j} f d\mu(\tau)$. For example for $f = 1$ we get that $V_\Gamma = [\mathrm{SL}_2(\mathbb{Z}) : \{\pm I\}\Gamma]V_{\mathrm{SL}_2(\mathbb{Z})}$. We are now ready to define the Petersson inner product.

Definition 2.3.14. *We define the Petersson inner product as $\langle .,.\rangle_\Gamma : \mathcal{S}_k(\Gamma) \times \mathcal{S}_k(\Gamma) \to \mathbb{C}$ with $\langle f, g\rangle_\Gamma = \frac{1}{V_\Gamma}\int_{X(\Gamma)} f(\tau)\overline{g(\tau)}(Im(\tau))^k d\mu(\tau)$.*

One can prove as in p.183 of [3.3] that this is indeed convergent and well defined and then a couple of other properties immediately emerge by the definition. In particular the Petersson inner product is:

(i) linear in f,

(ii) conjugate linear in g,

(iii) Hermitian symmetric,

(iv) positive definite.

The most useful property however is that the Petersson inner product is normal in the sense that it commutes with it's adjoint. In particular we can compute the adjoint operators for the Hecke algebra we defined earlier. Before that we need to define the inverse of a Diamond operator.

Definition 2.3.15. *Let $a \in \mathrm{SL}_2(\mathbb{Z})$ be any a such that $\langle d\rangle f = f[a]_k$. Then denote $\langle d\rangle^{-1}$ as the operator $\langle d\rangle f = f[a^{-1}]_k$. It follows that $\langle d\rangle\langle d\rangle^{-1} = \langle d\rangle^{-1}\langle d\rangle = 1$.*

Proposition 2.3.16. *Consider $\mathcal{S}_k(\Gamma_1(N))$ as an inner product space with the Petersson product. Then the adjoint of $\langle p\rangle$ is $\langle p\rangle^{-1}$ and the adjoint of T_p is $\langle p\rangle^{-1}T_p$ assuming $p \nmid N$. In particular both $\langle n\rangle$ and T_p are normal for $gcd(n, N) = 1$.*

Now we can use the Spectral Theorem of linear algebra on a finite dimensional inner product space with the Hecke algebra as a family of commuting, normal operators to obtain that:

Theorem 2.3.17. *The space of cusp forms $\mathcal{S}_k(\Gamma_1(N))$ has an orthonormal basis of simultaneous eigenforms for all Hecke operators $\{\langle n\rangle, T_n, gcd(n, N) = 1\}$.*

2.3d Oldforms and Newforms

We will now decompose the space of cuspforms into two spaces: a space of functions that arise from a form of lower level and brought up to a higher one with the use of an operator and those that are "genuinely new" and appear for the very first time at this specific level. Observe that $M|N$ means $\mathcal{S}_k(\Gamma_1(M)) \subseteq \mathcal{S}_k(\Gamma_1(N))$. We can move from $\mathcal{S}_k(\Gamma_1(M))$ to $\mathcal{S}_k(\Gamma_1(N))$ by applying the multiplication by d map which is simply $a_d = \begin{bmatrix} d & 0 \\ 0 & 1 \end{bmatrix}$ thus allowing us to define $f[a_d]_k(\tau) = d^{k-1}f(d\tau)$. This map is an injection $\mathcal{S}_k(\Gamma_1(M)) \to \mathcal{S}_k(\Gamma_1(N))$. We will try to distinguish the forms that come from lower levels.

Definition 2.3.18. *Define the map $i_d : \mathcal{S}_k(\Gamma_1(N/d))\times\mathcal{S}_k(\Gamma_1(N/d)) \to \mathcal{S}_k(\Gamma_1(N))$ with $i_d(f, g) = f+g[a_d]_k$, the space of oldforms of level N then is $\mathcal{S}_k(\Gamma_1(N))^{old} = \sum_{\substack{p|N \\ prime}} i_p(\mathcal{S}_k(\Gamma_1(N/p)), \mathcal{S}_k(\Gamma_1(N/p)))$ and the space of newforms $\mathcal{S}_k(\Gamma_1(N))^{new} = (\mathcal{S}_k(\Gamma_1(N))^{old})^{\perp}$ where the complement is taken with respect to the Petersson inner product.*

Quite notably the Hecke algebra preserves these spaces:

Proposition 2.3.19. *The spaces $\mathcal{S}_k(\Gamma_1(N))^{old}, \mathcal{S}_k(\Gamma_1(N))^{new}$ are stable under the Hecke operators $\{\langle n\rangle, T_n, gcd(n, N) = 1\}$ and both have orthonormal bases that are simultaneous eigenvectors for all these elements. Specifically $\mathcal{S}_k(\Gamma_1(N))^{new}$ has such a basis even when $gcd(n, N) > 1$.*

We will call an element $f \in \mathcal{M}_k(\Gamma_1(N))$ that is an eigenvector for all elements of the Hecke algebra an eigenform. A newform is a normalized eigenform ($a_1 = 1$ in the Fourier expansion) in $\mathcal{S}_k(\Gamma_1(N))^{new}$. Since $a_1(T_nf) = c_na_1(f)$ where c_n is the corresponding eigenvalue such that $T_nf = c_nf$, we have that $a_n(f) = c_na_1(f)$ away from the level.

Consider now a normalized version of $[a_d]$. Let $\iota_d = d^{1-k}[a_d]_k$. Then we have that $(\iota_d)f(\tau) = f(d\tau)$ and thus an action on the Fourier series as $\iota_d : \sum_{n=1}^{\infty} a_nq^n \to \sum_{n=1}^{\infty} a_nq^{dn}$. Notice that if $f = \sum_{p|N} \iota_p(f_p)$ with $f_p \in \mathcal{S}_k(\Gamma_1(N/p))$ then $gcd(n, N) = 1 \iff a_n = 0$. A notable result of Atkin-Lehner asserts that the opposite direction is also valid:

Theorem 2.3.20. ***(Atkin-Lehner)*** *If $f \in \mathcal{S}_k(\Gamma_1(N))$ has Fourier expansion $f(\tau) = \sum_{n=1}^{\infty} a_nq^n$ and $a_n = 0$ whenever $gcd(n, N) = 1$ then $f = \sum_{p|N} \iota_p(f_p)$ with $f_p \in \mathcal{S}_k(\Gamma_1(N/p))$.*

In the case of $a_1(f) = 0$ implies that $a_n = 0$ away from the level and the theorem above implies that $f \in \mathcal{S}_k(\Gamma_1(N))^{old}$. The above discussion indicates that the eigenvalues of the Hecke algebra define these spaces completely.

Theorem 2.3.21. *The set of newforms in the space $\mathcal{S}_k(\Gamma_1(N))^{new}$ form an orthogonal basis. Each such newform lies in an eigenspace $\mathcal{S}_k(N, \chi)$ and satisfies $a_n(f) = c_n a_1(f), \forall n \in \mathbb{N}$. Furthermore if $f, g \in \mathcal{S}_k(\Gamma_1(N))^{new}$ are both eigenforms with the same eigenvalues then they are equal up to a scalar multiple $f = \lambda g$.*

Also based on the above characterization of the Fourier coefficients of an eigenform we can further say that:

Proposition 2.3.22. *Let $f \in \mathcal{M}_k(N, \chi)$, then f is an eigenform if and only if:*

(i) $a_1(f) = 1$,

(ii) $a_{p^r}(f) = a_p(f)a_{p^{r-1}}(f) - \chi(p)p^{k-1}a_{p^{r-2}}(f)$,

(iii) $a_{mn}(f) = a_m(f)a_n(f)$ *whenever* $gcd(m, n) = 1$.

2.4 L-functions associated to modular curves

Let $f \in \mathcal{S}_k(\Gamma_1(N))$ with $f(\tau) = \sum_{n=1}^{\infty} a_n(f)q^n$. We define the corresponding L series as $L(s, f) = \sum_{n=1}^{\infty} \frac{a_n(f)}{n^s}$. In the case of cusp forms it can be shown that this series converges for all $s \in \mathbb{C} : Re(s) > k/2 + 1$. In the case of an eigenform we can also consider this as an Euler product by employing Prop. 2.3.22:

Proposition 2.4.1. *Let f be a normalized eigenform then $L(s, f) = \prod_p (1 - a_p p^{-s} + \chi(p)p^{k-1-2s})^{-1}$. The inverse also holds.*

A way to analytically extend an L series is with a functional equation that allows us to "mirror" it's behaviour on all of the complex plane. The way this is achieved is with a Mellin transform.

Definition 2.4.2. *Let $f \in \mathcal{S}_k(\Gamma_1(N))$, the Mellin transform of f is defined as $g(s) = \int_{t=0}^{\infty} f(it)t^{s-1}dt$ for $s : L(s, f)$ converges absolutely.*

The particular form of the Mellin transform of f in our case is:

$$g(s) = (2\pi)^{-s}\Gamma(s)L(s, f)$$

and we also set $\Lambda_N(s) = N^{s/2}g(s)$. We will find a functional equation for this Λ_N.

Consider the operator $W_N : \mathcal{S}_k(\Gamma_1(N)) \to \mathcal{S}_k(\Gamma_1(N))$ with $(W_N f)(\tau) = i^k N^{-k/2}\tau^{-k} f(-1/(N\tau))$. This operator is self adjoint and idempotent. Defining $\mathcal{S}_k(\Gamma_1(N))^{\pm} = \{f \in \mathcal{S}_k(\Gamma_1(N)) : W_n f = \pm f\}$ gives an orthogonal decomposition $\mathcal{S}_k(\Gamma_1(N)) = \mathcal{S}_k(\Gamma_1(N))^{+} \oplus \mathcal{S}_k(\Gamma_1(N))^{-}$. For cuspforms in these spaces it is now possible to determine a functional equation:

Theorem 2.4.3. *Suppose $f \in \mathcal{S}_k(\Gamma_1(N))$. Then the function Λ_N extends to an entire function satisfying the functional equation: $\Lambda_N(s) = \pm\Lambda_N(k - s)$. This implies $L(s, f)$ has an analytic continuation on all of $\mathbb{C}$.*

If we assume the Modularity Theorem then this is in fact the result we mentioned in Theorem 1.7.4.

2.5 Jacobians

In this section we will use geometric tools to associate to a newform a corresponding abelian variety. Consider a compact Riemann surface X of genus g and view it as a sphere with g handles. Label the loops around each handle as A_i for the longtitudinal loops and B_i for the latitudinal ones. The first homology group of X is then $H_1(X,\mathbb{Z}) = \sum_{i=1}^{g} a_i \int_{A_i} + b_i \int_{B_i} \cong \mathbb{Z}^{2g}$. The first homology group is a subgroup of $\Omega^1_{hol}(X)^* = \mathrm{Hom}_{\mathbb{C}}(\Omega^1_{hol}(X), \mathbb{C})$, the dual space of holomorphic differentials on X. The Jacobian is then defined as:

$$J(X) = \Omega^1_{hol}(X)^* / H_1(X,\mathbb{Z})$$

and it is in fact a natural definition if we consider $\Omega^1_{hol}(X)^*$ as integrating on X and want the integral to be independent from the specific path and to only depend on the points. Now consider again the exact same definitions for the divisors of an abelian variety as in the case of elliptic curves. In more detail the divisor group is the free abelian group generated by the points of X. We can then consider again $\mathrm{Pic}^0(X) = \mathrm{Div}^0(X)/\sim$ where the equivalence relation is for functions on X this time. X can always be embedded in the Jacobian via sending $x \to x - x_0/\sim$ where x_0 is a base point in X. Recall that we saw that in elliptic curves with the base point being O and the map $P \to P - O$. There is also a map $\mathrm{Div}^0(X) \to J(X)$ given by $\sum_{x\in X} n_x x \to \sum_{x\in X} n_x \int_{x_0}^{x}$ and it is well defined since the integrals depend only on the finite points $x \in X$ such that $n_x \neq 0$. In fact we have the following:

Theorem 2.5.1. *(**Abel's Theorem**) The map* $\mathrm{Pic}^0(X) \to J(X)$ *given by* $\sum_{x\in X} n_x x/\sim \to \sum_{x\in X} n_x \int_{x_0}^{x}$ *is an isomorphism.*

We thus embedded X in it's Picard group and by the theorem above in it's Jacobian with $X \to J(X)$ defined as $x \to \int_{x_0}^{x}$. By using the isomorphism in Abel's theorem we also see that $\Omega_1^*(X) = \{\sum_\gamma n_\gamma \int_\gamma : \sum_\gamma = 0\}$ where γ is a path in X.

2.5a Maps between Jacobians

Let $h : X \to Y$ be a nonconstant holomorphic map of Riemann surfaces and denote by $h^* : K(Y) \to K(X)$ the pullback on function fields.

Definition 2.5.2. *The forward map between Jacobians is the map* $h_J : J(X) \to J(Y)$ *given by* $h_J(\phi) = \phi \circ h^* / H_1(Y,\mathbb{Z})$.

In practice the effect of this map is $h_J : \sum_{x\in X} n_x \int_{x_0}^{x} \to \sum_{x\in X} n_x \int_{h(x_0)}^{h(x)}$. There is a corresponding map of Picard groups $h_P : \mathrm{Pic}^0(X) \to \mathrm{Pic}^0(Y)$ that commutes with h_J and Abel's isomorphism, namely the map $h_P : \sum_{x\in X} n_x x \to \sum_{x\in X} n_x h(x)$.

The other direction requires some more technical tools and specifically the trace. The first step is to remove all points of $x \in X, h(x) \in Y$ with $e_x > 1$ thus obtaining a $d-$fold covering map $h : X' \to Y'$. Let $\omega \in \Omega^1_{hol}(X)$ and suppose $y \in Y'$ so there exist local inverses $h_i^{-1} : U_y \to U_i, i \in \{1, ..., d\}$. Then the trace is defined as: $(tr_h \omega)_{U_y} = \sum_{i=1}^{d} (h_i^{-1})^* (\omega|_{U_i})$. The other direction is then:

Definition 2.5.3. *The reverse map of Jacobians* $h^J : J(Y) \to J(X)$ *is the holomorphic homomorphism* $h^J(\psi) = \psi \circ tr_h / H_1(X,\mathbb{Z})$ *where* $\psi \in \Omega^1_{hol}(Y)$.

The actual way this map acts on elements of the Jacobian is $h^J(\sum_{y\in Y} n_y \int_{y_0}^{y}) = \sum_{y\in Y} n_y \sum_{x\in h^{-1}(y)} e_x \int_{x_0}^{x}$. There is also a reverse Picard group homomorphism corresponding similarly to this map $h^P : \mathrm{Pic}^0(Y) \to \mathrm{Pic}^0(X)$ defined as $h(\sum_{y\in Y} n_y y) = \sum_{y\in Y} n_y \sum_{x\in h^{-1}(y)} e_x x$. We can show that $h_P \circ h^P$ is actually the same as multiplication by $deg(h)$ in $\mathrm{Pic}^0(Y)$ and thus by Abel's Theorem the same holds for $h_J \circ h^J$ in $J(Y)$. If this seems familiar it is because we have seen a specific example of this before: an elliptic curve isogeny and it's dual!

2.5b Jacobians of Modular Curves

We will now use Jacobians to formulate a geometric version of the Modularity Theorem for complex varieties. First we apply our results on Jacobians on modular curves. First denote $J_1(N) = J(X_1(N))$. Then the Hecke operators act naturally on $J_1(N)$. Going back to the double coset operators, let Γ_1, Γ_2 be congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$ with corresponding modular curves X_1, X_2. Then consider the configuration

$$\Gamma_2 \leftarrow a^{-1}\Gamma_1 a \cap \Gamma_2 \simeq \Gamma_1 \cap a\Gamma_2 a^{-1} \to \Gamma_1$$

which gives a corresponding configuration:

$$X_2 \overset{\pi_2}{\leftarrow} X(a^{-1}\Gamma_1 a \cap \Gamma_2) \overset{a}{\simeq} X(\Gamma_1 \cap a\Gamma_2 a^{-1}) \overset{\pi_1}{\to} X_1$$

with $\pi_2^{-1}(x) = \{e_y y : y \in X(a^{-1}\Gamma_1 a \cap \Gamma_2), \pi_2(y) = x\}$ and taking representatives $\Gamma_1 a \Gamma_2 = \cup_j \Gamma_1 \beta_j$ we get a map $\Gamma_2\tau \to \sum_j \Gamma_1 \beta_j \tau$ which can be linearly expanded to a map $[\Gamma_1 a \Gamma_2]_2 : \mathrm{Div}(X_2) \to \mathrm{Div}(X_1)$ and then it induces naturally a map on Picard groups $[\Gamma_1 a \Gamma_2]_2 : \mathrm{Pic}^0(X_2) \to \mathrm{Pic}^0(X_1)$. This is not just any map of divisors but a composition of forward and reverse maps with $[\Gamma_1 a \Gamma_2]_2 = (\pi_1)_P a_P (\pi_2)^P$. So how can we make sense of the differentials here? How will the geometry help us? Luckily we have the following:

Proposition 2.5.4. *The map $f(\tau) \to f(\tau)d\tau$ is an isomorphism $\mathcal{S}_2(\Gamma) \to \Omega^1_{hol}(X(\Gamma))$.*

This allows us to re-write $J(X(\Gamma)) = \mathcal{S}_2(\Gamma)^*/H_1(X(\Gamma), \mathbb{Z})$. Now we know that $[\Gamma_1 a \Gamma_2]_2 : \mathcal{S}_2(\Gamma_1) \to \mathcal{S}_2(\Gamma_2)$ is the map taking $f \to \sum_j f[\beta_j]_2$ which means it's pullback induces indeed a map on Jacobians taking $[\psi] \to [\psi \circ [\Gamma_1 a \Gamma_2]_2], \psi \in \mathcal{S}_2(\Gamma_2)^*$ where the brackets indicate up to homology equality. Setting $\Gamma_1 = \Gamma_2 = \Gamma_1(N)$ we can calculate the way a Hecke operator acts on Jacobians.

Proposition 2.5.5. *Let $T \in \{\langle d\rangle, T_p, d \in \mathbb{N}, p \text{ prime}\}$ be a Hecke operator on $\mathcal{S}_2(\Gamma_1(N))$. Then T acts on $J_1(N)$ as $T : J_1(N) \to J_1(N), [\phi] \to [\phi \circ T], \phi \in \mathcal{S}_2(\Gamma_1(N))^*$*

2.6 Algebraic Eigenvalues

In the last section we examined the way the Jacobian acts on modular curves and saw that the action of the Hecke operators $T : \mathcal{S}_2(\Gamma_1(N))^* \to \mathcal{S}_2(\Gamma_1(N))^*$ descends on the Jacobian. This means it must respect the kernel which is $H_1(X_1(N), \mathbb{Z})$ thus acting as an endomorphism on it. Let f be it's characteristic polynomial in the homology group, then $f(T_p) = 0$ in $H_1(X_1(N), \mathbb{Z})$

and $f \in \mathbb{Z}[X]$. By the $\mathbb{C}-$linearity of T_p we obtain that $f(T_p) = 0$ in the whole $\mathcal{S}_2(\Gamma_1(N))^*$ which means that if g is the characteristic polynomial of T_p in $\mathcal{S}_2(\Gamma_1(N))^*$ then $g|f \implies$ every root of $f \in \overline{\mathbb{Z}} \implies$ all eigenvalues of T_p are algebraic integers giving us that:

Proposition 2.6.1. *For $f \in \mathcal{S}_2(\Gamma_1(N))$ a normalized eigenform and every $n \in \mathbb{N}$ we have that $a_n(f)$ is an algebraic integer.*

We now look at the Hecke operators as an algebra over $\mathbb{Z}$ which we denote by $\mathbb{T}_\mathbb{Z}$. We can view this $\mathbb{Z}$-module as a ring of endomorphisms of the free and finitely generated $\mathbb{Z}$-module $H_1(X_1(N), \mathbb{Z})$ which immediately implies it is finitely generated. Define a map $\lambda_f : \mathbb{T}_\mathbb{Z} \to \mathbb{Z}$ taking $T \to \lambda_f(T)$ which is simply the eigenvalue of the normalized eigenform f and let $I_f = ker(\lambda_f) = \{T \in \mathbb{T}_\mathbb{Z} : Tf = 0\}$. We immediately see that

$$\mathbb{T}_\mathbb{Z}/I_f \simeq \mathbb{Z}[\{a_n(f)\}]$$

. Where the image is inside a number field K_f. The rank of $\mathbb{T}_\mathbb{Z}/I_f$ is then equal to the degree $[K_f : \mathbb{Q}]$.

Definition 2.6.2. *Let $f \in \mathcal{S}_k(\Gamma_1(N))$ be a normalized eigenform with $f(\tau) = \sum_{n=1}^{\infty} a_n(f)q^n$. The field $K_f = \mathbb{Q}(\{a_n(f)\})$ is called the number field of f.*

From what we saw above the name number field is indeed justified as it is a finite degree extension of $\mathbb{Q}$.

2.7 Abelian Varieties and Newforms

Start with a newform $f \in \mathcal{S}_k(\Gamma_1(N)$ which is as we already stated a normalized eigenfunction for $\mathbb{T}_\mathbb{Z}$. We established that $\mathbb{T}_\mathbb{Z}$ acts on $J_1(N)$ which means we have a subgroup $I_f J_1(N) \subseteq J_1(N)$.

Definition 2.7.1. *The abelian variety associated to f with notation as above is defined to be the quotient $A_f = J_1(N)/I_f J_1(N)$.*

With this definition $\mathbb{T}_\mathbb{Z}/I_f$ acts on A_f in a well defined way. Consider now the following equivalence relation on newforms: $f \sim f' \iff f' = f^\sigma$ for some automorphism $\sigma : \mathbb{C} \to \mathbb{C}$ and denote by $[f]$ the equivalence class of f. Define the space $V_f = \langle [f] \rangle \subseteq \mathcal{S}_k(\Gamma_1(N)$ meaning the space spanned by these newforms up to equivalence. Since $[f] = \{f^\sigma, \sigma \in K_f/\mathbb{Q}\}$ it has dimension equal to the degree $[K_f : \mathbb{Q}]$. Consider also $\Lambda_f = \mathrm{Hom}_1(X_1(N), \mathbb{Z})|_{V_f}$.

Proposition 2.7.2. *With notation as above there is an isomorphism $A_f \xrightarrow{\sim} V_f^*/\Lambda_f$ given by $[\phi] + I_f J_1(N) \to \phi_{V_f} + \Lambda_f$, where V_f^*/Λ_f is a complex torus of dimension $[K_f : \mathbb{Q}]$.*

We can extend the definition of an isogeny to complex tori of higher dimension.

Definition 2.7.3. *A surjective holomorphic homomorphism of complex tori with finite kernel is called an isogeny.*

We can now phrase the central theorem of this section:

Theorem 2.7.4. *The Jacobian associated to $\Gamma_1(N)$ is isogenous to a direct sum of abelian varieties associated to equivalence classes of newforms,*

$$J_1(N) \to \bigoplus_f A_f^{m_f}$$

where the sum is taken over a set of representatives $f \in \mathcal{S}_k(\Gamma_1(N_f))$ with m_f equal to the number of divisors of $N/N_f \in \mathbb{N}$.

We can now phrase the geometric version of the Modularity Theorem.

Theorem 2.7.5. ***(Modularity Theorem)*** *Let E be a complex elliptic curve with $j(E) \in \mathbb{Q}$. Then for some positive integer N there exists a newform $f \in \mathcal{S}_k(\Gamma_1(N))$ such that $A_f \to E$ is an isogeny.*

2.8 From Geometry to Algebra

In order to pass from geometry to number theory we will need to move from $\mathbb{C}$ to the integers and finite fields. To do so we will have to view modular curves from the perspective of algebraic geometry via their function fields. We will use the following definition form algebraic geometry:

Definition 2.8.1. *Let C be a non-singular affine algebraic curve over the rationals defined by polynomials $\phi_1, ..., \phi_m \in \mathbb{Z}_{(p)}[x_1, ..., x_n]$ where $\mathbb{Z}_{(p)}$ denotes the localization of $\mathbb{Z}$ over the prime p. Then C has good reduction if:*

(i) *the ideal $\langle \phi_1, ..., \phi_m \rangle$ is prime in $\mathbb{Z}_{(p)}[x_1, ..., x_n]$.*

(ii) *the reduced polynomials define a nonsingular affine algebraic curve over $\mathbb{F}_p$*

In the above definition the condition for a point to be singular is that the Jacobian matrix of the first derivatives at that point has rank lower than that of some other point of C. This means that C is nonsingular if and only if the Jacobian has constant rank at every point and so the above coincides with our usual definition in the case of an elliptic curve. Now it remains to see what happens with the morphism between reduced curves.

Theorem 2.8.2. *Let C, C' be nonsingular projective varieties over $\mathbb{Q}$ with good reduction at p and assume that C' has positive genus. Then the diagram*

$$\begin{array}{ccc} C & \xrightarrow{h} & C' \\ {\scriptstyle\sim}\downarrow & & \downarrow{\scriptstyle\sim} \\ \tilde{C} & \xrightarrow{\tilde{h}} & \tilde{C}' \end{array}$$

commutes and $deg(h) = deg(\tilde{h})$.

For divisors the situation is similar.

Theorem 2.8.3. *Let C be a nonsingular projective algebraic curve over $\mathbb{Q}$ with good reduction at p. The natural map induced on divisors by reduction $\sum n_p P \to \sum n_p \tilde{P}$ takes principal divisors to principal divisors and thus induces a bijection $\mathrm{Pic}^0(C) \to \mathrm{Pic}^0(\tilde{C})$. Moreover if C' satisfies the same conditions as C and has positive genus, then for every morphism $h : C \to C'$ we have that the diagram*

$$\begin{array}{ccc} \mathrm{Pic}^0(C) & \xrightarrow{h_*} & \mathrm{Pic}^0(C') \\ \sim\downarrow & & \downarrow\sim \\ \mathrm{Pic}^0(\tilde{C}) & \xrightarrow{\tilde{h_*}} & \mathrm{Pic}^0(\tilde{C'}) \end{array}$$

commutes, where h_* is the pushforward map of h on divisors.

This result is not obvious as the reduction map does not send $div(f) \to div(\tilde{f})$ in general. This theorem allows us to talk about reductions of Jacobians as we identified them naturally with the zero Picard group. We will next need a partial result of Igusa.

Theorem 2.8.4. *The modular curve $X_1(N)$ and it's Jacobian $J_1(N)$ have good reduction for all $p \nmid N$.*

This means that $X_1(N)$ defines a smooth algebraic curve over $\mathbb{F}_p$ and the corresponding Jacobian of this is the reduction of $J_1(N)$. We will now look at the reductions of the Hecke operators on reduced modular Jacobians.

The result is easy in the case of the $\langle d \rangle$ operator as Theorem 2.8.3 gives us a commutative diagram:

$$\begin{array}{ccc} J_1(N) & \xrightarrow{\langle d \rangle_*} & J_1(N) \\ \sim\downarrow & & \downarrow\sim \\ \tilde{J}_1(N) & \xrightarrow{\tilde{\langle d \rangle}_*} & \tilde{J}_1(N) \end{array}$$

The relation for T_p is given by the Eichler-Shimura relation. As $\tilde{X}_1(N)$ is defined in characteristic p, there is a natural morphism (the Frobenius endomorphism) defined as $\sigma_p : x \to x^p$. We then have the following.

Theorem 2.8.5. ***(Eichler-Shimura)*** *Let $p \nmid N$, then the following diagram commutes:*

$$\begin{array}{ccc} J_1(N) & \xrightarrow{T_p} & J_1(N) \\ \sim\downarrow & & \downarrow\sim \\ \tilde{J}_1(N) & \xrightarrow{\sigma_{p,*} + \tilde{\langle p \rangle}_* \sigma_p^*} & \tilde{J}_1(N) \end{array}$$

Notice that in $\tilde{X}_0(N)$ the operator $\tilde{\langle p \rangle}_*$ acts trivially and we thus get $\tilde{T}_p = \sigma_{p,*} + \sigma_p^*$ in this case. If this seems familiar again remember the result on elliptic curves where $[a_p(E)] = \sigma_p + \widehat{\sigma_p}$. Treating it as maps on divisors that would simply take the form $a_p(E) = \sigma_{p,*} + \sigma_p^*$ as an endomorphism on $\mathrm{Pic}^0(\tilde{E})$. This allows us to restate the geometric Modularity Theorem in a weaker but more computationaly accessible form:

Theorem 2.8.6. *Let E be an elliptic curve over $\mathbb{Q}$ with conductor N_E and $a : X_0(N) \to E$ be a rational morphism, then there exists $f \in \mathcal{S}_k(\Gamma_0(N_f))$ such that $N_f | N$ and $a_p(f) = a_p(E)$, for all primes $p \nmid N_E N$.*

Proof. (Sketch) By Theorem 2.7.5 there exists $A_f \to E$ with $A_f \subseteq J_0(N)$ and $a_p(f)$ is T_p as a map. When $p \nmid N$ we have good reduction implying

$\tilde{T}_p = \sigma_{p,*} + \sigma_p^*$. By Theorem 2.8.2 since $X_0(N), E$ have good reduction at p, we have that $\tilde{a}_*$ exists and commutes with $\tilde{T}_p$ giving us $\sigma_{p,*} + \sigma_p^*$ on $\mathrm{Pic}^0(\tilde{E})$ which is just as we saw above $a_p(E)$. □

The following stronger version also holds:

Theorem 2.8.7. *(**Modularity Theorem (L-functions)**) Let $E(\mathbb{Q})$ be an elliptic curve with conductor N. Then there exists $f \in \mathcal{S}_k(\Gamma_0(N))$ such that $L(s, f) = L(s, E)$.*

The use of Falting's Isogeny Theorem indicates that in fact this abelian variety A_f with an isogeny $A_f \to E$ we obtained from the geometric version is an elliptic curve.

2.9 Representations and Modularity

In this section we will showcase the connection between Galois representations and modular forms in a way similar to what we did for elliptic curves. Let N be a positive integer and ℓ be a prime. The modular curve $X_1(N)$ is a projective non-singular algebraic curve of genus g. The Jacobian is an abelian variety which means we can generalize our constructions on elliptic curves. Similar to the genus 1 case, there is an inclusion of torsion $i_n : J_1(N)[\ell^n]_{\mathbb{Q}} \to J_1(N)[\ell^n]_{\mathbb{C}} \simeq (\mathbb{Z}/\ell\mathbb{Z})^{2g}$ which is actually an isomorphism. As we already saw, Igusa's Theorem implies that the Jacobian $J_1(N)$ has good reduction for all $p \nmid N$, which gives a surjection $\pi_n : J_1(N)[\ell^n] \to \tilde{J}_1(N)[\ell^n]$. The map π_n is also an isomorphism. We can now generalize the construction of the Tate module.

Definition 2.9.1. *The ℓ-adic Tate module of $X_1(N)$ is $T_\ell(J_1(N)) = \varprojlim_n J_1(N)[\ell^n]$ and picking a basis we get that $T_\ell(J_1(N)) \simeq {\mathbb{Z}_\ell}^{2g}$.*

The Galois group $G_{\mathbb{Q}}$ acts on divisors in the natural way:

$$(\sum n_P P)^\sigma \to \sum n_P P^\sigma$$

and since $(div(f))^\sigma = div(f^\sigma)$ this action descends to the Picard group and in partcular it repsects the map $[\ell] : J_1(N)[\ell^{n+1}] \to J_1(N)[\ell^n]$. This means we have a continuous representation

$$\rho_{X_1(N),\ell} : G_{\mathbb{Q}} \to \mathrm{GL}_{2g}(\mathbb{Z}_\ell)$$

which is the Galois representation associated to $X_1(N)$. The following result is similar to the one we proved for elliptic curves but instead we use the Eichler-Shimura relation on T_p.

Theorem 2.9.2. *Let ℓ be a prime and N a positive integer. The Galois representation $\rho_{X_1(N),\ell}$ is then unramified at all primes $p \nmid \ell N$ and for any such p if $\mathfrak{p}$ is a maximal ideal of $\mathbb{Z}$ lying over p, then the characteristic polynomial of $\rho_{X_1(N),\ell}(Frob_{\mathfrak{p}})$ is $x^2 - T_p x + \langle p \rangle p$. Moreover the Galois representation is irreducible.*

So far we saw the situation for $X_1(N)$ but what about a newform $f \in \mathcal{S}_2(N,\chi)$? The geometric object linked to a newform in that case is $A_f \simeq J_1(N)/I_f J_1(N)$ where $I_f = \{T \in \mathbb{T}_{\mathbb{Z}} : Tf = 0\}$. The dimension of A_f as a complex torus is $d = [K_f : \mathbb{Q}]$. We also have an isomorphism $\mathbb{T}_{\mathbb{Z}}/I_f \simeq \mathbb{Z}[\{a_n(f) : n \in \mathbb{N}\}] = \mathcal{O}_f$. With this assumption the action of $a_p(f)$ on A_f is $T_p + I_f$ and the action of $\chi(p)$ is $\langle p\rangle + I_f$. We now define the Tate module for a newform.

Definition 2.9.3. *Let $f \in \mathcal{S}_2(N,\chi)$, the ℓ-adic Tate module of f is then $T_\ell(A_f) = \varprojlim_n A_f[\ell^n]$ and picking a basis we get that $T_\ell(A_f) \simeq {\mathbb{Z}_\ell}^{2d}$.*

Now we have the following useful fact:

Lemma 2.9.4. *The map $J_1(N)[\ell^n] \to A_f[\ell^n]$ is a surjection and it's kernel is stable under $G_{\mathbb{Q}}$.*

Proof. The multiplication map $[\ell^n] : x \to \ell^n x$ is surjective on the d-torus which implies it is also surjective on $I_f J_1(N)$ since $y \in I_f J_1(N) \implies y = \sum T_i y_i, T_i \in I_f, y_i \in J_1(N)$ and thus $y_i = \ell^n x_i, x_i \in J_1(N)$ which means $y = \ell^n \sum T_i x_i$. The kernel is easily seen to be $I_f J_1(N)[\ell^n]$ which is stable under the absolute Galois group as the actions of Hecke operators commute with the Galois action on $J_1(N)$. □

This means that $G_{\mathbb{Q}}$ acts on $A_f[\ell^n]$ and thus on $T_\ell(A_f)$. Since it commutes with the action of the Hecke operators on $T_\ell(A_f)$, it also commutes with the action of $\mathcal{O}_f$. We thus obtain a Galois representation:

$$\rho_{A_f,\ell} : G_{\mathbb{Q}} \to \mathrm{GL}_{2d}(\mathbb{Z}_\ell)$$

which can be shown to be continuous. We denote $\rho_{f,\ell} = \rho_{A_f,\ell}$. The dimension of this representation is $2d$ but we need a representation of dimension 2 in order to associate an elliptic curve to our newform f. This is achieved via the lemma below.

Lemma 2.9.5. *The tensor product $T_\ell(A_f) \otimes \mathbb{Q}_\ell$ is a free module of rank 2 over $K_f \otimes \mathbb{Q}_\ell$.*

Using the fact that $K_f \otimes \mathbb{Q}_\ell = \prod_{\lambda|\ell} K_{f,\lambda}$ and projecting gives us a representation

$$\rho_{f,\lambda} : G_{\mathbb{Q}} \to \mathrm{GL}_2(K_{f,\lambda})$$

which can also be shown to be continuous and $ker(\rho_{f,\ell}) \subseteq \rho_{f,\lambda}$. Since we identified the action T_p as $a_p(f)$ and of $\langle p\rangle$ as $\chi(p)$ we have the following theorem.

Theorem 2.9.6. *Let ℓ be a prime and $f \in \mathcal{S}_2(N,\chi)$ a normalized eigenform with number field K_f. Then for each maximal ideal $\lambda \supseteq \ell$ in $\mathcal{O}_{K_f}$ there exists a 2-dimensional Galois representation*

$$\rho_{f,\lambda} : G_{\mathbb{Q}} \to \mathrm{GL}_2(K_{f,\lambda})$$

that is unramified at every prime $p \nmid \ell N$. For any such p if $\mathfrak{p}$ is a maximal ideal of $\mathbb{Z}$ lying over p, then the characteristic polynomial of $\rho_{f,\ell}(Frob_{\mathfrak{p}})$ is $x^2 - a_p(f)x + \chi(p)p$. In particular if $f \in \mathcal{S}_2(\Gamma_0(N))$ then the characteristic polynomial is $x^2 - a_p(f)x + p$.

With this striking similarity between the Galois representations attached to an elliptic curve and to a modular form we will restate the modularity theorem in a representation theoretic version.

Theorem 2.9.7. *(**Modularity Theorem for Representations**) Let E be an elliptic curve over $\mathbb{Q}$ with conductor N. Then there is $f \in \mathcal{S}_2(\Gamma_0(N))$ with rational number field such that $\rho_{E,\ell} \sim \rho_{f,\ell}$ for all primes ℓ.*

Chapter 3

The Modular Approach

The modular approach is one of the strongest strategies in the study of Diophantine equations. An almost trivial example of the method that showcases this strength is Wiles' proof of Fermat's Lasts Theorem. In it's core it is a technique based on contradiction: namely we attach a Frey curve to a supposed solution of a Diophantine equation, associate it to a space of newforms and either show why none of these fit our original requirements or we obtain a bound for one of our unkown parameters.

Before we begin to showcase the modular approach we note that some deep results required will have to be taken on trust. This chapter will end up being more computational and to fully utilise the method we will use SageMath.

3.1 Level Lowering

In this section we will present a collection of definitions and notable results on Ribet's Theorem (see [3.3]) which we will use in our approach. First of all, every newform in this chapter is of weight $k = 2$. When we say "the newforms of level N" we mean a normalized (and as we already saw finite) eigenbasis for $\mathcal{S}_2^{new}(N)$.

Definition 3.1.1. *Let E be a rational elliptic curve and $f = \sum_{n\geq 1} c_n q^n$ be a newform with associated number field $K_f/\mathbb{Q}$. We say that E arises from the newform f and write $E \overset{p}{\sim} f$ if there is a prime ideal $\mathfrak{p}|p$ of K_f such that $\bar{\rho}_{E,\mathfrak{p}} \sim \bar{\rho}_{f,\mathfrak{p}}$.* $\mod p$

In particular taking traces and applying the above relation to the Forbenius element $Frob_\ell$ we get that $a_\ell(E) = c_\ell \mod \mathfrak{p}$ for almost all prime ℓ. We can get more precise however. One can in fact guess that the primes for which this relation fails to hold is when one of the representations is ramified. Let N_E, N_f be the conductors of E and f and use Theorems 1.8.7 and 2.9.6 to conclude the following.

Corollary 3.1.2. *Suppose $E \overset{p}{\sim} f$, then there is a prime ideal $\mathfrak{p}|p$ of K_f such that for every prime ℓ:*

(i) *if $\ell \nmid pN_EN_f$ then $a_\ell(E) \equiv c_\ell \mod \mathfrak{p}$*

(ii) *if $\ell \nmid pN_f$ and $\ell \| N_E$ then $\ell + 1 \equiv \pm c_\ell \mod \mathfrak{p}$.*

In the case of a rational newform ($K_f = \mathbb{Q}$) there is a rational elliptic curve F that is associated to f by the Modularity Theorem. We then say that $E \overset{p}{\sim} F$. In that case a result of Kraus and Oestrelé (see [3.3]) allows us to get rid of the unkown p in the condition:

Proposition 3.1.3. *If E, F are rational elliptic curves with conductors N_E, N_F such that $E \overset{p}{\sim} F$ then for every prime ℓ:*

(i) *if $\ell \nmid N_E N_F$ then $a_\ell(E) \equiv a_\ell(F) \mod \mathfrak{p}$*

(ii) *if $\ell \nmid N_F$ and $\ell \| N_E$ then $\ell + 1 \equiv \pm a_\ell(F) \mod \mathfrak{p}$.*

The first condition corresponds to the case when both elliptic curves have good reduction at ℓ whereas the second amounts to the case of F having good reduction and E multiplicative reduction at ℓ. We now define the following constant associated to an elliptic curve E with conductor N and minimal discriminant Δ_{min} and a prime p:

$$N_p = N / \prod_{\substack{q \| N \\ p | ord_q(\Delta_{min})}} q,$$

which is basically the conductor without the primes that are raised to a p-th power in the minimal discriminant. We will use a simplified case of Ribet's Level Lowering Theorem:

Theorem 3.1.4. ***(Ribet)*** *Suppose E is a rational elliptic curve with no p-isogenies, where $p \geq 3$ prime. Let N_p be as above, then there exists a newform f such that $E \overset{p}{\sim} f$ and f has level N_p.*

Ribet's Theorem is a very strong tool: it introduces a $\mod \mathfrak{p}$ relation on one hand and sacrifices the rationality condition for the newform but it allows us to get rid of the part of the conductor that comes from the unkown variables in our supposed solution that are raised to the p-th power. This makes it feasible to test all possible newform spaces in some cases where N_p only has constants in it like in FLT.

We thus observe that in order to apply Ribet's Theorem we must first show that the curve in question has no p-isogenies. There are a couple of ways to do this.

Definition 3.1.5. *Let $E/\mathbb{Q}$ be an elliptic curve such that for every prime p, E has either good or multiplicative reduction at p. We call such an elliptic curve semi-stable.*

If the conductor is squarefree, then the asociated elliptic curve is always semi-stable. The following Theorem is a special case of Mazur's work (see [3.3]):

Theorem 3.1.6. *Let $E/\mathbb{Q}$ be a semi-stable elliptic curve with $E[2] = 4$ and $p \geq 5$ prime. Then E has no p-isogenies.*

The main Theorem we will be using however is the following.

Theorem 3.1.7. *Suppose $E/\mathbb{Q}$ is an elliptic curve with conductor N such that $ord_2(N) = 3, 5$ or 7. Then E has no isogenies of odd degree.*

3.2 The Modular Approach

With everything settled we are finally able to describe the Modular Approach. We will first start with an example, namely a direct application on Fermat's Last Theorem (for Wile's original paper see [3.3]).

3.2a Fermat's Last Theorem

Suppose we had a non trivial solution of the equation:

$$a^n + b^n = c^n$$

for some prime $n \geq 5$. Consider now the elliptic curve:

$$E : Y^2 = X(X - a^n)(X + b^n) \tag{3.1}$$

which is called a Frey curve. A Frey curve is an elliptic curve arising from a hypothetical solution of a Diophantine equation. Using Tate's algorithm as presented in [3.3] we compute the minimal discriminant and the conductor of E:

$$\Delta_{min} = 2^{-8}(abc)^{2n}, \;\; N = 2 \prod_{\substack{p|abc \\ p \neq 2}} p,$$

as well as the conductor used in level lowering (noting that one of a, b, c has to be even):

$$N_n = 2.$$

Theorem 3.1.6 implies that the curve (3.1) has no n-isogenies and thus we can use Ribet's Theorem to conclude that there exists a level 2 newform f such that $E \overset{n}{\sim} f$. Note however that from Proposition 2.5.4 we have that $dim(S_2(\Gamma(2))) = genus(X(2))$ and using Theorem 2.2.4 one can show this genus is 0 (see ex. 3.1.4 [3.3]), meaning there can be no such newform f.

3.2b Generalizing the Method

The immediate question that emerges now is whether we can use this method to tackle other similar problems and what structure should these problems have in order for this approach to be useful? The most common type of such problems in the area of Diophantine equations are problems where we have an equation involving unkown exponents such that we can construct a Frey curve out of a supposed solution. This curve needs to have a suitable discriminant and some other properties in order for Ribet's Level Lowering theorem to work. In case we get a non-singular curve notice that we do not have an elliptic curve which is the case for the trivial solution of the FLT for example. So in general we need the following conditions to be met:

(i) A Frey curve with coefficients depending on a solution,

(ii) A minimal discriminant of the form $\Delta_{min} = CD^n$ where n is the unkown exponent and C depends only on the equation and not on the solution,

(iii) E has multiplicative reduction at all primes $p|D$ meaning that when we take N_n then $p \nmid N_n$.

After that finitely many newform of level N_n for which it is possible to have $E \overset{n}{\sim} f$. Now this leaves us with two cases:

(i) If indeed $a_p(E) = c_p$ for every prime p then we have to find a way to eliminate this possibility or it could mean a solution actually exists.

(ii) If $a_p(E) \neq c_p$ then recall that $a_p(E) \equiv c_p \mod \mathfrak{n}$ where $\mathfrak{n}$ is a prime ideal over n in K_f and thus $n = N(\mathfrak{n}) | N(a_p(E) - c_p) \neq 0$ which means we have a bound for the unkown exponent!

Below we will discuss applications of this method to instances of a specific Diophantine problem.

3.3 Applications to Diophantine Equations

Consider the equation:

$$x^2 + d^2 = 2y^n, \tag{3.2}$$

where we assume x, y, d to be pairwise coprime. Notice that the coprimality assumption forces $d, y = 1 \mod 2$ and similarly to the FLT case we can assume without loss of generality that $n \geq 5$ is prime. We will use the signature recipes in 13.2 of [3.3] to produce the Frey curve after we bring it to the suitable form:

$$2y^n + d^2(-1)^n = x^2,$$

which upon inspection leads us to case (ii) which is the Frey curve:

$$E(x, y) : Y^2 = X^3 + 2xX + 2y^n X,$$

provided of coure that $xy \neq 0$. This is a perfect showcase of the first limitation of the Modular Approach as presented here: in case $d = 1$ we get the equation:

$$x^2 + 1 = 2y^n,$$

with the obvious solution $x = y = 1$ for every $n \in \mathbb{N}$. This means there is actually a newform attached to our Frey curve in this case and thus any attempt at a contradiction argument will always fail. We will use again Tate's algorithm to compute the invariants of this curve:

$$D_{min} = -2^8 d^2 y^{2n} \quad , \quad N = 2^7 \operatorname{rad}(d) \operatorname{rad}(y)$$

Observing that $ord_2(N) = 7$ we can use Theorem 3.1.7 to conclude that our Frey curve has no n-isogeny. Ribet's Theorem then allows us to attach to this curve a newform of level:

$$N_n = 2^7 \operatorname{rad}(d)$$

which depends on d.

We will now focus on the cases $d = 3, 7$ and bound the exponent n. In order to do that we will use SageMath to efficiently implement the following algorithm:

(i) Set $N_n = 2^7 \operatorname{rad}(d)$ define the parametric curve $E(x, y)$ with $\Delta_y = 2^8 d y^{2n}$,

(ii) find the newforms of level N_n and pick a specific newform f,

(iii) pick from the first m primes $P = \{p_1, ..., p_m\}$ such that $p_i \nmid N_n, \forall i$,

(iv) for each prime $p \in P$ compute $B = \prod_{a,b \mod p} (a_p(E_{a,b}) - a_p(f))$, whenever $E_{a,b}$ is non singular and $a^2 + d^2 \equiv b \mod p$,

(v) compute $B_p = \begin{cases} ((p+1)^2 - a_p(f)^2)B, & f \text{ rational} \\ p((p+1)^2 - a_p(f)^2)B, & f \text{ irrational} \end{cases}$, to account for the second case of Corollary 3.1.2 and the fact that we could have $p = n$,

(vi) Compute $d_f = gcd(B_{p_1}, ..., B_{p_m})$ and note that $n | d_f$

(vii) output the largest prime factor of d_f

The corresponding SageMath program is shown below:

```
def newform_elimination(d,bound):
    Ex = lambda x, y: EllipticCurve([0, 2*x, 0, 2*y, 0])
    N_F = factor(d).radical_value()*2**7
    newforms = Newforms(N_F, names='a')
    newfsGCD = []
    for fnew in newforms:
        Bps = []
        for p in [p for p in range(3, bound) if p in Primes()]:
            if N_F%p!= 0:
                Bp = 1
                apfnew = fnew[p]
                for a in range(p):
                    for b in range(p):
                        dEx = 2*b*d
                        if dEx%p != 0 and (a**2 + d**2 - 2*b)%p ==
    0:
                            w = (apfnew - (p + 1 - Ex(a, b).
    reduction(p).order()))
                            Bp *= (apfnew - (p + 1 - Ex(a, b).
    reduction(p).order()))
                Bp *= ((p+1)**2 - apfnew**2)
                if fnew.base_ring() is not QQ:
                    Bp *= p
                Bps.append(ZZ(Bp.norm()))
        GCD = gcd(Bps)
        if not GCD.is_zero():
            if not GCD == 1:
                newfsGCD.append(max(GCD.prime_factors()))
            else:
                newfsGCD.append(1)
        else:
            newfsGCD.append(0)
    print("newform bounds =", newfsGCD)
```

In the case $d = 3$ and $m = 100$ we get the following result:

```
d = 3
bound = 100
newform_elimination(d,bound)
```

Output

```
newform bounds = [5, 3, 3, 5, 5, 3, 3, 5]
```

The above means that using the Modular Approach we can conclude that if $d = 3$ then equation 3.2 can only have a solution in the case $n \leq 5$.

Similarly for $d = 7$ and $m = 100$:

```
d = 7
bound = 100
newform_elimination(d,bound)
```

Output

```
newform bounds = [3, 3, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2]
```

So in the case $d = 7$ we get an even lower bound of $n \leq 3$.

Bibliography

[1] M. J. Silverman, *The Arithmetic of Elliptic Curves*, 2nd edn. Springer, 2008.

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

[4] F. Diamond, J.Shurman *A First Course in Modular Forms*, Springer Science+Business Media 2005.

[5] Kraus, Oestrelé, *Sur une question de B. Mazur*, Math. Ann. 293(1992), p. 259–275.

[6] B.Mazur, *Rational isogenies of prime degree*, Invent. Math. 44(1978), p. 129–162.

[7] S.Siksek, *The Modular Approach to Diophantine Equations*, in Explicit Methods in Number Theory: Rational Points and Diophantine Equations, Panoramas et synthèses 36 (2012)

[8] K. Ribet, *On modular representations of* $Gal(\bar{\mathbb{Q}}\mathbb{Q})$ *arising from modular curves*, Invent. Math. 100(1990), p. 431-476.

[9] A. Wiles, *Modular elliptic curves and Fermat's Last Theorem*, Annals of Math. 141(1995), p. 443-551.