Hecke Algebras and Modular Forms:
Notes derived from Ribet's 1996 Berkeley grad. course.

William Stein

September 24, 1998

Preface

Disclaimer: These notes record some of what I saw in Ken Ribet's course on Modular Forms and Hecke Operators given at U.C. Berkeley during the Spring semester 1996. They are still *very rough* as I wrote them during my first semester of graduate school before I knew any real mathematics.

The participants in the course were: Amod Agashe, Matt Baker, Jim Borger, Kevin Buzzard, Bruce Caskel, Robert Coleman, Janós Csirik, Annette Huber, David Jones, David Kohel, Loic Merel, David Moulton, Andrew Ogg, Arthur Ogus, Jessica Polito, Ken Ribet, Saul Schleimer, Lawren Smithline, William Stein, Takahashi, Wayne Whitney, and Hui Zui.

I wish to thank David Moulton, Joe Wetherall, and Kevin Buzzard who helped me in preparing these notes, Arthur Ogus who asked a lot of stimulating questions during the class, and of course Ken Ribet who sees clearly.

William Stein, Spring 1996, Berkeley, CA, `was@math.berkeley.edu`

Contents

Preface **iii**

1 Introduction **1**
1.1 Two Dimensional Galois Representations . . . 1
1.1.1 Finite Fields (Weil, Tate) . . . 1
1.1.2 Galois Representations (Taniyama, Shimura, Mumford-Tate) . . . 2
1.2 Modular Forms and Galois Representations . . . 2
1.2.1 Cusp Forms . . . 2
1.2.2 Hecke Operators (Mordell) . . . 2

2 Modular Representations and Curves **5**
2.1 Arithmetic of Modular Forms . . . 5
2.2 Characters . . . 6
2.3 Parity Conditions . . . 6
2.4 Conjectures of Serre (mod ℓ version) . . . 7
2.5 General remarks on mod p Galois representations . . . 7
2.6 Serre's Conjecture . . . 8
2.7 Wiles' Perspective . . . 8

3 Modular Forms **9**
3.1 Cusp Forms . . . 9
3.2 Lattices . . . 9
3.3 Relationship With Elliptic Curves . . . 9
3.4 Hecke Operators . . . 10
3.5 Explicit Description of Sublattices . . . 11
3.6 Action of Hecke Operators on Modular Forms . . . 12

4 Embedding Hecke Operators in the Dual **15**
4.1 The Space of Modular Forms . . . 15
4.2 Inner Product . . . 16
4.3 Eigenforms . . . 17

5 Rationality and Integrality Questions **19**
5.1 Review . . . 19
5.2 Integrality . . . 19
5.3 Victor Miller's Thesis . . . 20
5.4 Petersson Inner Product . . . 20

6 Modular Curves **23**
6.1 Cusp Forms . . . 23
6.2 Modular Curves . . . 23
6.3 Classifying $\Gamma(N)$-structures . . . 24
6.4 More on Integral Hecke Operators . . . 24
6.5 Complex Conjugation . . . 25
6.6 Isomorphism in the Real Case . . . 25
6.7 The Eichler-Shimura Isomorphism . . . 25
6.8 The Petterson Inner Product is Hecke Compatible . . . 27

7 Higher Weight Modular Forms **29**
7.1 Definitions of $\mathbf{T}$ . . . 29
7.2 Double Cosets . . . 29
7.3 More General Congruence Subgroups . . . 30
7.4 Explicit Formulas . . . 31
7.5 Old and New Forms . . . 31

8 New Forms **33**
8.1 Connection With Galois Representations . . . 34
8.2 Semisimplicity of U_p . . . 34
8.3 Shimura's Example of Nonsemisimple U_p . . . 34
8.4 An Interesting Duality . . . 35
8.5 Observations on T_n . . . 36

9 Some Explicit Genus Computations **37**
9.1 Computing the Dimension of $S_k(\Gamma)$ . . . 37
9.2 Application of Riemann-Hurwitz . . . 37
9.3 Explicit Genus Computations . . . 38
9.4 The Genus of $X(N)$ . . . 38
9.5 The Genus of $X_0(N)$ . . . 39
9.6 Modular Forms mod p . . . 40

10 The Field of Moduli **41**
10.1 Digression on Moduli . . . 41
10.2 When is ρ_E Surjective? . . . 42
10.3 Observations . . . 43
10.4 A Descent Problem . . . 44
10.5 Second Look at the Descent Exercise . . . 44
10.6 Action of GL_2 . . . 45

11 Hecke Operators as Correspondences **47**
11.1 Some Philosophy . . . 47
11.2 Hecke Operators as Correspondences . . . 48
11.3 Generalities on Correspondences . . . 49
11.4 Jacobians of Curves . . . 50
11.5 More on Hecke Operators . . . 51
11.6 Hecke Operators acting on Jacobians . . . 51
11.6.1 The Albanese Map . . . 52
11.6.2 The Hecke Algebra . . . 53
11.7 The Eichler-Shimura Relation: Part I . . . 53

11.8 The Eichler-Shimura Relation: Part II . . . 54
11.9 Applications . . . 56
11.10 More on Eichler-Shimura . . . 57

12 Abelian Varieties from Modular Forms **59**
12.1 Computing the Determinent of ρ_λ . . . 61
12.2 Duality and Polarizations . . . 62
12.3 The Weil Pairing . . . 63
12.4 The Fancy Proof . . . 63
12.5 The Concrete Proof . . . 64
12.6 The Construction for $X_1(N)$ . . . 64

13 The Gorenstein Property **67**
13.1 The Gorenstein Property . . . 69
13.2 Proof the Gorenstein Property . . . 72
13.2.1 Vague Comments . . . 75
13.3 Finite Flat Group Schemes . . . 75
13.4 Reformulation of $V = W$ problem . . . 75
13.5 Dieudonné Theory . . . 76
13.6 The Proof: Part II . . . 77
13.7 Key Result of Boston-Lenstra-Ribet . . . 79

14 Local Properties of ρ_λ **81**
14.1 Definitions . . . 81
14.2 Local Properties when $p\nmid N$ . . . 81
14.3 Weil-Deligne Groups . . . 82
14.4 Local Properties when $p|N$ . . . 82
14.5 Definition of the Reduced Conductor . . . 83
14.6 Introduction . . . 84
14.7 Adelic Representations Associated to Modular Forms . . . 84
14.8 More Local Properties of the ρ_λ. . . . 87
14.8.1 Possibilities for π_p . . . 87
14.8.2 The case $\ell = p$ . . . 88
14.8.3 Tate Curves . . . 89

15 The Weight and Serre's Conjectures **91**
15.1 Introduction . . . 91
15.2 Review of the λ-adic case . . . 91
15.3 Serre's conjecture 0 . . . 91
15.3.1 Problems . . . 92
15.4 Serre's conjecture 1 . . . 92
15.4.1 Key background points . . . 93
15.5 The weight and fundamental characters . . . 94
15.6 The weight in Serre's conjectures on modular representations . . . 98
15.6.1 θ-series . . . 98
15.6.2 Edixhoven's paper . . . 100
15.7 The extra assumption . . . 100
15.7.1 Companion Forms . . . 102
15.8 The exceptional level 1 case . . . 103

16 Fermat's Last Theorem **105**
16.1 The application to Fermat . . . 105
16.2 Modular Elliptic Curves . . . 106

17 Deformations **109**
17.1 Introduction . . . 109
17.2 Condition $(*)$ . . . 110
17.2.1 Finite flat representations . . . 111
17.3 Classes of Liftings . . . 111
17.3.1 The case $p \neq \ell$ . . . 111
17.3.2 The case $p = \ell$ . . . 112
17.4 Wiles' Hecke algebra . . . 112

18 The Hecke Algebra T_Σ **115**
18.1 The Hecke Algebra . . . 115
18.2 The maximal ideal in R . . . 117
18.2.1 Strip away certain Euler factors . . . 117
18.2.2 Make into an eigenform for U_ℓ . . . 118
18.3 The Galois Representation . . . 118
18.3.1 The structure of $\mathbf{T}_\mathbf{m}$ . . . 120
18.3.2 The philosophy in this picture . . . 120
18.3.3 Massage ρ . . . 120
18.3.4 Massage ρ' . . . 121
18.3.5 Representations from modular forms mod ℓ . . . 121
18.3.6 Representations from modular forms mod ℓ^n . . . 122
18.4 ρ' is of type Σ . . . 122
18.5 Isomorphism between $\mathbf{T}_\mathbf{m}$ and $R_{\mathbf{m}_R}$ . . . 123
18.6 Deformations . . . 124
18.7 Wiles Main Conjecture . . . 125
18.8 $\mathbf{T}_\Sigma$ is a complete intersection . . . 127
18.9 The inequality $\#\mathcal{O}/\eta \leq \#\wp_T/\wp_T^2 \leq \wp_R/\wp_R^2$ . . . 127
18.9.1 The definitions of the ideals . . . 128
18.9.2 Aside: Selmer Groups . . . 129
18.9.3 Outline of some proofs . . . 129

Chapter 1

Introduction

The main objects of study in this course are:

- Modular Forms
- Hecke Algebras
- Modular Curves
- Jacobians
- Abelian Varieties

1.1 Two Dimensional Galois Representations

The geometric objects of study are elliptic curves and more generally algebraic curves of arbitrary genus. These in turn give rise via the Jacobian construction to higher dimensional abelian varieties. These geometric objects in turn give rise to Galois representations.

When studying elliptic curves, the natural tool in the characteristic zero situation is to present the elliptic curve as $\mathbf{C}/\mathcal{L}$ for some lattice $\mathcal{L}$ in $\mathbf{C}$. To construct $\mathcal{L}$ fix a non-zero holomorphic differential ω of E over $\mathbf{C}$ and construct $\mathcal{L}$ as

$$\left\{ \int_\gamma \omega \;\middle|\; \gamma \in H_1(E(\mathbf{C}), \mathbf{Z}) \right\}.$$

1.1.1 Finite Fields (Weil, Tate)

In the 1940's, Weil study the analogous situation for elliptic curves defined over a finite field k. He desperately wanted to find an algebraic way to describe the above correspondence. He was able to find an algebraic definition of $\mathcal{L}/n\mathcal{L}$, where $n \geq 1$ and $(n, \operatorname{char} k) = 1$, which is as follows. Let $E[n] = \{P \in E(\overline{k}) : nP = 0\} = (\frac{1}{n}\mathcal{L})/\mathcal{L} \cong \mathcal{L}/n\mathcal{L}$.

Now fix a prime ℓ, we let $E[\ell^\infty] = \{P \in E(\overline{k}) : \ell^\nu P = 0, \text{ some } \nu \geq 1\} = \cup_{\nu=1}^{\infty} E[\ell^\nu]$. Tate obtained an analogous construction by defining a rank 2 free $\mathbf{Z}_\ell$-module $\operatorname{Tate}_\ell E := \varprojlim E[\ell^\nu]$ (the map from $E[\ell^\nu] \to E[\ell^{\nu-1}]$ is multiplication by ℓ). To see that the rank is 2, check that the $\mathbf{Z}/\ell^\nu\mathbf{Z}$-module structure of $E[\ell^\nu]$ is compatible with the maps $E[\ell^\nu] \to E[\ell^{\nu-1}]$). See [33] (III, 7). Then $V_\ell(E) = T_\ell(E) \otimes \mathbf{Q}_\ell$ is a two dimensional vector space over $\mathbf{Q}_\ell$. This gives the first nontrivial example of ℓ-adic étale cohomology.

1.1.2 Galois Representations (Taniyama, Shimura, Mumford-Tate)

Let $E/\mathbf{Q}$ be an elliptic curve and $G = \mathcal{G}al(\overline{\mathbf{Q}}/\mathbf{Q})$. Then $E[n] = \{P \in E(\overline{\mathbf{Q}}) : nP = 0\} \cong (\mathbf{Z}/n\mathbf{Z})^2$ is acted on by G and this action respects the group operation so we have a Galois representation

$$G \xrightarrow{\rho} \mathrm{Aut}(E[n]) \cong \mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z})$$

Let K be the fixed field of $\ker\rho$ (note that K is a number field), then since $\mathcal{G}al(K/\mathbf{Q}) \cong G/\ker\rho \cong \mathrm{Im}\,\rho \subseteq \mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z})$ we obtain many subgroups of $\mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z})$ as Galois groups. Shimura showed that if we start with the elliptic curve

$$E: \quad y^2 + y = x^3 - x^2$$

then the image of ρ is often all of $\mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z})$ and the image is "most" of $\mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z})$ when E does not have complex multiplication.

1.2 Modular Forms and Galois Representations

1.2.1 Cusp Forms

Let $S_k(N)$ denote the space of cusp forms of weight k and level N on the congruence subgroup $\Gamma_1(N) = \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z}) : a \equiv 1 \pmod{N}, c \equiv 0 \pmod{N}, d \equiv 1 \pmod{N}\right\}$. Thus $S_k(N)$ is the finite dimensional vector space consisting of all holomorphic functions $f(z)$ on $\mathcal{H} = \{z \in \mathbf{C} : \mathrm{Im}(z) > 0\}$ vanishing at ∞ and satisfying

$$f(\frac{az+b}{cz+d}) = (cz+d)^k f(z) \quad \text{for all} \quad \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N).$$

Since, in particular, $f(z) = f(z+1)$ we can expand $f(z)$ as a q-series (this requires rigorous justification)

$$f(z) = \sum_{n=1}^{\infty} c_n q^n.$$

A famous example is

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

τ is called the Ramanujan function. One now knows that τ is multiplicative and satisfies $\tau(p^\nu) = \tau(p)\tau(p^\nu) - p^{11}\tau(p^{\nu-1})$. Δ is a normalized basis for $S_1(1)$.

1.2.2 Hecke Operators (Mordell)

Mordell defined, for $n \geq 1$, operators T_n on $S_k(N)$ called *Hecke operators*. These proved very fruitful. The set of such operators forms an "almost" commuting family of endomorphisms and is hence "almost" simultaneously diagonalizable. The precise meaning of "almost" and the actual structure of the Hecke algebra $\mathbf{Q}[T_1, \ldots, T_n, \ldots]$ will be studied in greater detail in the remainder of this course. Often there will exist a basis of cusp forms $f = \sum_{n=1}^{\infty} c_n q^n \in S_k(N)$ so that f_n is a simultaneous eigenvector for all of the Hecke operators T_n and, in fact,

$T_n f = c_n f$. All of the c_n will be algebraic integers and the field $\mathbf{Q}(c_1, c_2, \dots)$ will be finite over $\mathbf{Q}$.

A good claim can be made that the c_n are often interesting integers because they exhibit remarkable properties. For example, $\tau(n) \equiv \sum_{d|n} d^{11} \pmod{691}$. How can we study the c_n? How can we interpret the c_n? We can do this by studying the connection between Galois representations and modular forms. In 1968 work was originally begun on this by Serre, Shimura, Eichler and Deligne.

Chapter 2

Modular Representations and Curves

2.1 Arithmetic of Modular Forms

Suppose $f = \sum_{n=1}^{\infty} a_n q^n$ is a cusp form in $S_k(N)$ which is an eigenform for the Hecke operators. Then the Mellin transform associates to f the L-function $L(f, z) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$. Let $K = \mathbf{Q}(a_1, a_2, \ldots)$, then one can show that the a_n are algebraic integers and K is a number field. When $k = 2$ Shimura associates to f an abelian variety A_f over $\mathbf{Q}$ of dimension $[K : \mathbf{Q}]$ on which K acts (see theorem 7.14 of [31]).

Example 2.1.1 (Modular Elliptic Curves). When all of the coefficients a_n of the modular form f lie in $\mathbf{Q}$ then $[K : \mathbf{Q}] = 1$ so A_f is a one dimensional abelian variety. A one dimensional abelian variety of nonzero genus is an elliptic curve. An elliptic curve isogenous to one arising via this construction is called *modular*.

Definition 2.1.2. Elliptic curves E_1 and E_2 are *isogenous* if there is a morphism $E_1 \to E_2$ of algebraic groups, which has a finite kernel.

The following conjecture motivates much of the theory.

Conjecture 2.1.3. Every elliptic curve over $\mathbf{Q}$ is modular, that is, isogenous to a curve constructed in the above way.

For $k \geq 2$ Serre and Deligne found a way to associate to f a family of ℓ-adic representations. Let ℓ be a prime number and K be as above, then it is well known that $K \otimes_{\mathbf{Q}} \mathbf{Q}_\ell \cong \prod_{\lambda | \ell} K_\lambda$. One can associate to f a representation

$$\rho_{\ell,f} : G = \mathcal{G}al(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(K \otimes_{\mathbf{Q}} \mathbf{Q}_\ell)$$

unramified at all primes $p \not| \ell N$. For $\rho_{\ell,f}$ to be unramified we mean that for all primes P lying over p, the inertia group of the decomposition group at P is contained in the kernel of $\rho_{\ell,f}$. The decomposition group D_P at P is the set of those $g \in G$ which fix P. Let k be the residue field $\mathcal{O}/P$ where $\mathcal{O}$ is the ring of all algebraic integers. Then the inertia group I_P is the kernel of the map $D_P \to \mathcal{G}al(\overline{k}/k)$.

Now $I_P \subset D_P \subset \mathcal{G}al(\overline{\mathbf{Q}}/\mathbf{Q})$ and D_P/I_P is cyclic (being isomorphic to a subgroup of the Galois group of a finite extension of finite fields) so it is generated by a Frobenious automorphism frob_p lying over p. One has

$$\mathrm{Tr}(\rho_{\ell,f}(\mathrm{frob}_p)) = a_p \in K \subset K \otimes \mathbf{Q}_\ell$$

and

$$\det(\rho_{\ell,f}) = \chi_\ell^{k-1} \varepsilon$$

where χ_ℓ is the ℓth cyclotomic character and ε is the Dirichlet character associated to f. There is an incredible amount of "abuse of notation" packed into this statement. First, the Frobenius frob_P (note P not p) is only well defined in $\mathcal{G}al(K/\mathbf{Q})$ (so I think an unstated result is that K must be Galois), and then frob_p is only well defined up to conjugacy. But this works out since $\rho_{\ell,f}$ is well-defined on $\mathcal{G}al(K/\mathbf{Q})$ (it kills $\mathcal{G}al(\overline{\mathbf{Q}}/K)$) and the trace is well-defined on conjugacy classes ($\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ so $\mathrm{Tr}(ABA^{-1}) = Tr(B)$).

2.2 Characters

Let $f \in S_k(N)$, then for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ with $c \equiv 0 \mod N$ we have

$$f(\frac{az+b}{cz+d}) = (cz+d)^k \varepsilon(d) f(z)$$

where $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ is a Dirichlet character mod N. If f is an eigenform for the so called "diamond-bracket operator" $\langle d \rangle$ so that $f|\langle d \rangle = \varepsilon(d) f$ then ε actually takes values in K.

Led φ_N be the mod N cyclotomic character so that $\varphi_N : G \to (\mathbf{Z}/N\mathbf{Z})^*$ takes $g \in G$ to the automorphism induced by g on the Nth cyclotomic extension $\mathbf{Q}(\boldsymbol{\mu}_N)$ of $\mathbf{Q}$ (where we identify $\mathcal{G}al(\mathbf{Q}(\boldsymbol{\mu}_N)/\mathbf{Q})$ with $(\mathbf{Z}/N\mathbf{Z})^*$). Then what we called ε above in the formula $\det(\rho_\ell) = \chi_\ell^{k-1}\varepsilon$ is really the composition

$$G \xrightarrow{\varphi_N} (\mathbf{Z}/N\mathbf{Z})^* \xrightarrow{\varepsilon} \mathbf{C}^*.$$

For each positive integer ν we consider the ℓ^νth cyclotomic character on G,

$$\varphi_{\ell^\nu} : G \to (\mathbf{Z}/\ell^\nu\mathbf{Z})^*.$$

Putting these together gives the ℓ-adic cyclotomic character

$$\chi_\ell : G \to \mathbf{Z}_\ell^*.$$

2.3 Parity Conditions

Let $c \in \mathcal{G}al(\overline{\mathbf{Q}}/\mathbf{Q})$ be complex conjugation. Then $\varphi_N(c) = -1$ so $\varepsilon(c) = \varepsilon(-1)$ and $\chi_\ell^{k-1}(c) = (-1)^{k-1}$. Now let $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, then for $f \in S_k(N)$,

$$f(z) = (-1)^k \varepsilon(-1) f(z)$$

so $(-1)^k\varepsilon(-1) = 1$ thus

$$\det(\rho_{\ell,f}(c)) = \epsilon(-1)(-1)^{k-1} = -1.$$

Thus the det character is odd so the representation $\rho_{\ell,f}$ is odd.

Remark 2.3.1 (Vague Question). How can one recognize representations like $\rho_{\ell,f}$ "in nature"? Mazur and Fontaine have made relevant conjectures. The Shimura-Taniyama conjecture can be reformulated by saying that for any representation $\rho_{\ell,E}$ comming from an elliptic curve E there is f so that $\rho_{\ell,E} \cong \rho_{\ell,f}$.

2.4 Conjectures of Serre (mod ℓ version)

Suppose f is a modular form, $\ell \in \mathbf{Z}$ prime, λ a prime lying over ℓ, and the representation

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

(constructed by Serre-Deligne) is irreducible. Then $\rho_{\lambda,f}$ is conjugate to a representation with image in $\mathrm{GL}_2(\mathcal{O}_\lambda)$, where $\mathcal{O}_\lambda$ is the ring of integers of K_λ. Reducing mod λ gives a representation

$$\overline{\rho}_{\lambda,f} : G \to \mathrm{GL}_2(\mathbf{F}_\lambda)$$

which has a well-defined trace and det, i.e., the det and trace don't depend on the choice of conjugate representation used to obtain the reduced representation. One knows from representation theory that if such a representation is semisimple then it is completely determined by its trace and det (more precisely, the characteristic polynomials of all of its elements – see chapter ??). Thus if $\overline{\rho}_{\lambda,f}$ is irreducible (and hence semisimple) then it is unique in the sense that it does not depend on the choice of conjugate.

2.5 General remarks on mod p Galois representations

[[This section was written by Joseph Loebach Wetherell.]]

First, what are semi-simple and irreducible representations? Remember that a representation ρ is a map from a group G to the endomorphisms of some vector space W (or a free module M if we are working over a ring instead of a field, but let's not worry about that for now). A subspace W' of W is said to be invariant under ρ if ρ takes W' back into itself. (The point is that if W' is invariant, then ρ induces representations on both W' and W/W'.) An irreducible representation is one where the only invariant subspaces are 0 and W. A semi-simple representation is one where for every invariant subspace W' there is a complementary invariant subspace W'' – that is, you can write ρ as the direct sum of $\rho|_{W'}$ and $\rho|_{W''}$.

Another way to say this is that if W' is an invariant subspace then we get a short exact sequence

$$0 \to \rho|_{W/W'} \to \rho \to \rho|_{W'} \to 0.$$

Furthermore ρ is semi-simple if and only if every such sequence splits.

Note that irreducible representations are semi-simple.

One other fact is that semi-simple Galois representations are uniquely determined (up to isomorphism class) by their trace and determinant.

Now, since in the case we are doing, $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is compact, it follows that the image of any Galois representation ρ into $\mathrm{GL}_2(K_\lambda)$ is compact. Thus we can conjugate it into $\mathrm{GL}_2(\mathcal{O}_\lambda)$. Irreducibility is not needed for this.

Now that we have a representation into $\mathrm{GL}_2(\mathcal{O}_\lambda)$, we can reduce to get a representation $\overline{\rho}$ to $\mathrm{GL}_2(\mathbf{F}_\lambda)$. This reduced representation is not uniquely determined by ρ, since we had a choice of conjugators. However, the trace and determinant are invariant under conjugation, so the trace and determinant of the reduced representation are uniquely determined by ρ.

So we know the trace and determinant of the reduced representation. If we also knew that it was semi-simple, then we would know its isomorphism class, and we would be done. So we would be happy if the reduced representation is irreducible. And in fact, it is easy to see that if the reduced representation is irreducible, then ρ must also be irreducible. Now, it turns out that all ρ of interest to us will be irreducible; unfortunately, we can't go the other way and claim that ρ irreducible implies the reduction is irreducible.

2.6 Serre's Conjecture

Serre has made the following conjecture which is still open at the time of this writing.

Conjecture 2.6.1 (Serre). All irreducible representation of G over a finite field which are odd, i.e., $det(\sigma(c)) = -1$, c complex conjugation, are of the form $\overline{\rho}_{\lambda,f}$ for some representation $\rho_{\lambda,f}$ constructed as above.

Example 2.6.2. Let $E/\mathbf{Q}$ be an elliptic curve and let $\sigma_\ell : G \to \mathrm{GL}_2(\mathbf{F}_\ell)$ be the representation induced by the action of G on the ℓ-torsion of E. Then $\det \sigma_\ell = \varphi_\ell$ is odd and σ_ℓ is usually irreducible, so Serre's conjecture would imply that σ_ℓ is modular. From this one can, assuming Serre's conjecture, prove that E is modular.

Definition 2.6.3. Let $\sigma : G \to \mathrm{GL}_2(\mathbf{F})$ ($\mathbf{F}$ is a finite field) be a represenation of the Galois group G. The we say that the *represention σ is modular* if there is a modular form f, a prime λ, and an embedding $\mathbf{F} \hookrightarrow \overline{\mathbf{F}}_\lambda$ such that $\sigma \cong \overline{\rho}_{\lambda,f}$ over $\overline{\mathbf{F}}_\lambda$.

2.7 Wiles' Perspective

Suppose $E/\mathbf{Q}$ is an elliptic curve and $\rho_{\ell,E} : G \to \mathrm{GL}_2(\mathbf{Z}_\ell)$ the associated ℓ-adic representation on the Tate module T_ℓ. Then by reducing we obtain a mod ℓ representation

$$\overline{\rho}_{\ell,E} = \sigma_{\ell,E} : G \to \mathrm{GL}_2(\mathbf{F}_\ell).$$

If we can show this representation is modular for infinitely many ℓ then we will know that E is modular.

Theorem 2.7.1 (Langland's and Tunnel). *If $\sigma_{2,E}$ and $\sigma_{3,E}$ are irreducible, then they are modular.*

This is proved by using the fact that $\mathrm{GL}_2(\mathbf{F}_2)$ and $\mathrm{GL}_2(\mathbf{F}_3)$ are solvable so we may apply "base-change".

Theorem 2.7.2 (Wiles). *If ρ is an ℓ-adic representation which is irreducible and modular mod ℓ with $\ell > 2$ and certain other reasonable hypothesis are satisfied, then ρ itself is modular.*

Chapter 3

Modular Forms

Our goal is to explain modular forms as functions of lattices or of elliptic curves. Good references are Serre [24] and Katz [1].

3.1 Cusp Forms

First suppose $N = 1$, then we must define $S_k = S_k(1)$. Let $\Gamma_1(1) = \mathrm{SL}_2(\mathbf{Z})$, then S_k consists of all functions f holomorphic on the upper half plane $\mathcal{H}$ and such that for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$ one has

$$f(\frac{a\tau + b}{c\tau + d}) = (c\tau + d)^k f(\tau),$$

and f vanishes at infinity. Thus, in particular, $f(\tau + 1) = f(\tau)$ and so f passes to a well defined function of $q = e^{2\pi i\tau}$. So $f(q)$ is a function on $\{z : 0 < |z| < 1\}$ and the condition that $f(\tau)$ vanishes at infinity is that $f(q)$ extends to a holomorphic function on $\{z : |z| < 1\}$ and $f(0) = 0$. In this case, we may write $f(q) = \sum_{n=1}^{\infty} a_n q^n$.

3.2 Lattices

A lattice $L \subset \mathbf{C}$ is a subring $L = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$ for which $\omega_1, \omega_2 \in \mathbf{C}$ are lineary independent over $\mathbf{R}$. Without loss, we may assume that $\omega_1/\omega_2 \in \mathcal{H}$.

Let

$$\mathcal{R} = \{\text{lattices in } \mathbf{C}\} = \{(E, \omega) : \ E \text{ is an elliptic curve, } \omega \in \Omega^1_E\}$$

and

$$M = \{(\omega_1, \omega_2) : \omega_1, \omega_2 \in \mathbf{C}, \mathrm{Im}(\omega_1/\omega_2) > 0\}.$$

There is a left action of $\mathrm{SL}_2(\mathbf{Z})$ on M

$$\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) : (\omega_1, \omega_2) \mapsto (a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)$$

and $\mathrm{SL}_2(\mathbf{Z})\backslash M \cong \mathcal{R}$.

3.3 Relationship With Elliptic Curves

There is a map $L \mapsto \mathbf{C}/L$ from lattices to complex tori which, by Weierstrass theory, correspond to elliptic curves defined over $\mathbf{C}$ along with a distinguished differential $\omega = dz$.

Conversely, if $E/\mathbf{C}$ is an elliptic curve, we can obtain the corresponding lattice by fixing a differential ω and taking the lattice to be the image of the map

$$H_1(E(\mathbf{C}),\mathbf{Z}) \xrightarrow{\text{integration}} \mathbf{C}$$

which takes $\gamma \in H_1$ to $\int_\gamma \omega \in \mathbf{C}$.

There is a map $M/\mathbf{C} \to \mathcal{H}$ defined by $(\omega_1,\omega_2) \mapsto \omega_1/\omega_2$. This gives an isomorphism

$$\mathcal{R}/\mathbf{C}^* = (\mathrm{SL}_2(\mathbf{Z})\backslash M)/\mathbf{C}^* \to \mathrm{SL}_2(\mathbf{Z})\backslash\mathcal{H}$$

and

$$\mathcal{R}/\mathbf{C}^* = \{\text{ elliptic curves } /\mathbf{C} \text{ (without differentials)}\}.$$

If $f:\mathcal{H}\to\mathbf{C}$ we define $F: M \to \mathbf{C}$ by $F(\omega_1,\omega_2) = f(\omega_1/\omega_2)$. Suppose now that F is a lattice function and sattisfies the homogeneity condition $F(\lambda L) = \lambda^{-k}F(L)$. Then

$$\begin{aligned}
f\left(\frac{a\tau+b}{c\tau+d}\right) &= F\left(\mathbf{Z}\frac{a\tau+b}{c\tau+d}+\mathbf{Z}\right)\\
&= F((c\tau+d)^{-1}(\mathbf{Z}(a\tau+b)+\mathbf{Z}(c\tau+d)))\\
&= (c\tau+d)^k F(\mathbf{Z}(a\tau+b)+\mathbf{Z}(c\tau+d))\\
&= (c\tau+d)^k F(\mathbf{Z}+\tau\mathbf{Z})\\
&= (c\tau+d)^k f(\tau)
\end{aligned}$$

so functions of lattices with the homogeneity condition come from functions $f \in M_k$. Thus, if $f \in M_k$ and F is the corresponding lattice function then

$$F(\mathbf{Z}\omega_1+\mathbf{Z}\omega_2) = F\left(\omega_2\left(\mathbf{Z}+\mathbf{Z}\frac{\omega_1}{\omega_2}\right)\right) = \omega_2^{-k}F\left(\mathbf{Z}+\mathbf{Z}\frac{\omega_1}{\omega_2}\right) = \omega_2^{-k}f\left(\frac{\omega_1}{\omega_2}\right),$$

so we can recover F from f.

3.4 Hecke Operators

Define a map T_n from the free abelian group generated by all $\mathbf{C}$-lattices into itself by

$$T_n(L) = \sum_{(L:L')=n} L'.$$

Then if F is a function on lattices define T_nF by

$$(T_nF)(L) = n^{k-1}\sum_{(L:L')=n} F(L').$$

Since $(n,m)=1$ implies $T_nT_m = T_{nm}$ and T_{p^k} is a polynomial in $\mathbf{Z}[T_p]$ the essential case to consider is n prime.

Suppose $L' \subset L$ with $(L:L') = n$, then L/L' is killed by n so $nL \subset L' \subset L$ and

$$L'/nL \subset L/nL \cong (\mathbf{Z}/n\mathbf{Z})^2.$$

Thus the subgroups of $(\mathbf{Z}/n\mathbf{Z})^2$ of order n correspond to the sublattices L' of L of index n. When $n=\ell$ is prime there are $\ell+1$ such subgroups. (The subgroups correspond to nonzero vectors in $\mathbf{F}_\ell$ modulo scalar equivalence and there are $\frac{\ell^2-1}{\ell-1}$ of them.)

Suppose $L' \subset L$ is a sublattice of index ℓ and let $L'' = \ell^{-1}L'$. Note that $\ell L \subset L'$ so $L \subset \ell^{-1}L' = L''$ and L is a sublattice of L'' of index ℓ. Thus, assuming F satisfies the homogeneity condition $F(\lambda L) = \lambda^{-k}F(L)$,

$$\ell^{k-1}\sum_{L'} F(L') = \frac{1}{\ell}\sum_{L''} F(L'')$$

which helps explain the extra factor of n^{k-1} in our definition of T_nF – we are "averaging" over the sublattices (note that there are $\ell+1$ terms yet we divide by ℓ so we aren't exactly averaging).

We now give a geometric description of the ℓth Hecke operator. Let $L \subset L''$ be lattices with $(L'' : L) = \ell$ and let $E = \mathbf{C}/L$, $E'' = \mathbf{C}/L''$ be the elliptic curves corresponding to L, L'', respectively. Then $E[\ell] = \frac{1}{\ell}L/L$ contains $H = L''/L$ which may be thought of as a line [Ed: I don't know why!]. Then the Hecke operator is

$$E \mapsto \frac{1}{\ell}\sum_{\text{lines } H\subset E[\ell]} E/H.$$

Let $\hat{\pi}$ be the isogeny dual to $\pi : E \to E/H$. Then in terms of pairs (E,ω) we have

$$(E,\omega) \mapsto \frac{1}{\ell}\sum_{H\subset E[\ell],\#H=\ell} (E/H, \pi_*\omega) = \ell^{k-1}\sum_{H\subset E[\ell]} (E/H, \hat{\pi}^*(\omega)).$$

We consider modular forms f on $\Gamma_1(1) = \mathrm{SL}_2(\mathbf{Z})$, that is, holomorphic functions on $\mathcal{H}\cup\{\infty\}$ which satisfy

$$f(\tau) = f(\frac{a\tau+b}{c\tau+d})(c\tau+d)^{-k}$$

for all $\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$. Using a Fourier expansion we write

$$f(\tau) = \sum_{n=0}^{\infty} a_n e^{2\pi i\tau n},$$

and say f is a cusp form if $a_0 = 0$. There is a correspondence between modular forms f and lattice functions F satisfying $F(\lambda L) = \lambda^{-k}F(L)$ given by $F(\mathbf{Z}\tau + \mathbf{Z}) = f(\tau)$.

3.5 Explicit Description of Sublattices

The nth Hecke operator T_n of weight k is defined by

$$T_n(L) = n^{k-1}\sum_{\substack{L'\subset L\\(L:L')=n}} L'.$$

What are the L' explicitly? Note that L/L' is a group of order n and

$$L'/nL \subset L/nL = (\mathbf{Z}/n\mathbf{Z})^2.$$

Write $L = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$, let Y_2 be the cyclic subgroup of L/L' generated by ω_2 and let $d = \#Y_2$. Let $Y_1 = (L/L')/Y_2$, then Y_1 is generated by the image of ω_1 so it is a cyclic group of order $a = n/d$. We want to exhibit a basis of L'. Let $\omega_2' = d\omega_2 \in L'$ and use the fact that Y_1 is generated by ω_1 to write $a\omega_1 = \omega_1' + b\omega_2$ for some integer b and some $\omega_1' \in L'$. Since b is only well-defined modulo d we may assume $0 \le b \le d-1$. Thus

$$\begin{pmatrix}\omega_1'\\ \omega_2'\end{pmatrix} = \begin{pmatrix}a & b\\ 0 & d\end{pmatrix}\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}$$

and the change of basis matrix has determinent $ad = n$. Since

$$\mathbf{Z}\omega_1' + \mathbf{Z}\omega_2' \subset L' \subset L = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$$

and $(L : \mathbf{Z}\omega_1' + \mathbf{Z}\omega_2') = n$ (since the change of basis matrix has determinent n) and $(L : L') = n$ we see that $L' = \mathbf{Z}\omega_1' + \mathbf{Z}\omega_2'$.

Thus there is a one-to-one correspondence between sublattices $L' \subset L$ of index n and matrices $\left(\begin{smallmatrix}a & b\\ 0 & d\end{smallmatrix}\right)$ with $ad = n$ and $0 \le b \le d-1$. In particular, when $n = p$ is prime there $p+1$ of these. In general, the number of such sublattices equals the sum of the positive divisors of n.

3.6 Action of Hecke Operators on Modular Forms

Now assume $f(\tau) = \sum_{m=0}^{\infty} c_m q^m$ is a modular form with corresponding lattice function F. How can we describe the action of the Hecke operator T_n on $f(\tau)$? We have

$$\begin{aligned}
T_n F(\mathbf{Z}\tau + \mathbf{Z}) &= n^{k-1} \sum_{\substack{a,b,d\\ ab=n\\ 0\le b<d}} F((a\tau+b)\mathbf{Z} + d\mathbf{Z})\\
&= n^{k-1}\sum d^{-k} F(\frac{a\tau+b}{d}\mathbf{Z} + \mathbf{Z})\\
&= n^{k-1}\sum d^{-k} f(\frac{a\tau+b}{d})\\
&= n^{k-1}\sum_{a,d,b,m} d^{-k} c_m e^{2\pi i(\frac{a\tau+b}{d})m}\\
&= n^{k-1}\sum_{a,d,m} d^{1-k} c_m e^{\frac{2\pi i a m\tau}{d}} \frac{1}{d}\sum_{b=0}^{d-1}(e^{\frac{2\pi i m}{d}})^b\\
&= n^{k-1}\sum_{\substack{ad=n\\ m'\ge 0}} d^{1-k} c_{dm'} e^{2\pi i a m'\tau}\\
&= \sum_{\substack{ad=n\\ m'\ge 0}} a^{k-1} c_{dm'} q^{am'}.
\end{aligned}$$

In the second to the last expression we let $m = dm'$, $m' \ge 0$, then used the fact that the sum $\frac{1}{d}\sum_{b=0}^{d-1}(e^{\frac{2\pi i m}{d}})^b$ is only nonzero if $d|m$.

Thus

$$T_n f(q) = \sum_{\substack{ad=n\\ m\ge 0}} a^{k-1} c_{dm} q^{am}$$

and if $\mu \ge 0$ then the coefficient of q^μ is

$$\sum_{\substack{a|n\\ a|\mu}} a^{k-1} c_{\frac{n\mu}{a^2}}.$$

Remark 3.6.1. When $k \geq 1$ the coefficients of q^μ for all μ belong to the **Z**-module generated by the c_m.

Remark 3.6.2. Setting $\mu = 0$ gives the constant coefficient of $T_n f$ which is

$$\sum_{a|n} a^{k-1} c_0 = \sigma_{k-1}(n) c_0.$$

Thus if f is a cusp form so is $T_n f$. ($T_n f$ is holomorphic since its original definition is as a finite sum of holomorphic functions.)

Remark 3.6.3. Setting $\mu = 1$ shows that the coefficient of q in $T_n f$ is $\sum_{a|1} 1^{k-1} c_n = c_n$. As an immediate corollary we have the following important result.

Corollary 3.6.4. *Suppose f is a cusp form for which $T_n f$ has 0 as coefficient of q for all $n \geq 1$, then $f = 0$.*

Remark 3.6.5. When $n = p$ is prime we get an interesting formula for the action of T_p on the q-expansion of f. One has

$$T_p f = \sum_{\mu \geq 0} \sum_{\substack{a|n \\ a|\mu}} a^{k-1} c_{\frac{n\mu}{a^2}} q^\mu.$$

Since $n = p$ is prime either $a = 1$ or $a = p$. When $a = 1$, $c_{p\mu}$ occurs in the coefficient of q^μ and when $a = p$, we can write $\mu = p\lambda$ and we get terms $p^{k-1} c_\lambda$ in $q^{\lambda p}$. Thus

$$T_n f = \sum_{\mu \geq 0} c_{p\mu} q^\mu + p^{k-1} \sum_{\lambda \geq 0} c_\lambda q^{p\lambda}.$$

Chapter 4

Embedding Hecke Operators in the Dual

4.1 The Space of Modular Forms

Let $\Gamma = \Gamma_1(1) = \mathrm{SL}_2(\mathbf{Z})$ and for $k \geq 0$ let

$$\begin{aligned} M_k &= \{f = \sum_{n=0}^{\infty} a_n q^n : f \text{ is a modular form for } \Gamma\} \\ &\subset S_k = \{f = \sum_{n=1}^{\infty} a_n q^n\} \end{aligned}$$

These are finite dimensional $\mathbf{C}$-vector spaces whose dimensions are easily computed. Furthermore, they are generated by familiar elements (see Serre [24] or Lang [10].) The main tool is the formula

$$\sum_{p \in D \cup \{\infty\}} \frac{1}{e(p)} \operatorname{ord}_p(f) = \frac{k}{12}$$

where D is the fundamental domain for Γ and

$$e(p) = \begin{cases} 1 & \text{otherwise} \\ 2 & \text{if } p = i \\ 3 & \text{if } p = \rho \end{cases}$$

One can alternatively define $e(p)$ as follows. If $p = \tau$ and $E = \mathbf{C}/(\mathbf{Z}\tau + \mathbf{Z})$ then $e(p) = \frac{1}{2}\#\operatorname{Aut}(E)$.

For $k \geq 4$ we define the *Eisenstein series* G_k by

$$G_k(q) = \frac{1}{2}\zeta(1-k) + \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n,$$

then the map

$$\tau \mapsto \sum_{\substack{(m,n) \neq (0,0) \\ m,n \in \mathbf{Z}}} \frac{1}{(m\tau + n)^k}$$

differs from G_k by a constant (no proof). Also, $\zeta(1-k) \in \mathbf{Q}$ and one may say, *symbolically at least*, "$\zeta(1-k) = \sum_{d=1}^{\infty} d^{k-1} = \sigma_{k-1}(0)$." The *nth Bernoulli number* B_n is defined by the equation

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

One can show that $\zeta(1-k) = -\frac{B_k}{k}$ so the constant coefficient of G_k is $-\frac{B_k}{2k}$ which is rational.

4.2 Inner Product

In what follows we assume $k \geq 2$ to avoid trivialities.. The Hecke operators T_n acts on the space M_k. Fix a subspace $V \subset M_k$ which is stable under the action of the T_n. Let $\mathbf{T}(V)$ be the $\mathbf{C}$-algebra generated by the endomorphism T_n acting on V and note that $\mathbf{T}(V)$ is actually a finite dimensional $\mathbf{C}$-vector space since it is a subspace of $End(V)$ and V is finite dimensional. Recall that $\mathbf{T}$ is commutative.

There is a bilinear form

$$\begin{aligned} &\mathbf{T} \times V \to \mathbf{C} \\ &\langle T, f \rangle \mapsto a_1(f|T) \end{aligned}$$

where $f|T = \sum_{n=0}^{\infty} a_n(f|T) q^n$. We thus get maps

$$\begin{aligned} V &\to \mathrm{Hom}(\mathbf{T}, \mathbf{C}) = \mathbf{T}^* \\ \mathbf{T} &\to \mathrm{Hom}(V, \mathbf{C}) = V^*. \end{aligned}$$

Theorem 4.2.1. *The above maps are isomorphisms.*

Proof. It just remains to show each map is injective. Then since a finite dimensional vector space and its dual have the same dimension the result follows. First suppose $f \mapsto 0 \in \mathrm{Hom}(\mathbf{T}, \mathbf{C})$, then $a_1(f|T) = 0$ for all $T \in \mathbf{T}$ so, in particular, $a_n = a_1(f|T_n) = 0$ for all $n \geq 1$. Thus f is a constant, but since $k \geq 2$ this implies $f = 0$ (otherwise f wouldn't transform correctly with respect to the action of the modular group).

Next suppose $T \mapsto 0 \in \mathrm{Hom}(V, \mathbf{C})$, then $a_1(f|T) = 0$ for all $f \in V$. Substiting $f|T_n$ for f and using the commutativity of $\mathbf{T}$ we have

$$\begin{aligned} a_1((f|T_n)|T) &= 0 && \text{for all } f,\ n \geq 1 \\ a_1((f|T)|T_n) &= 0 && \text{by commutativity} \\ a_n(f|T) &= 0 && n \geq 1 \\ f|T &= 0 && \text{since } k \geq 2\text{, as above} \end{aligned}$$

Thus $T = 0$ which completes the proof. □

Remark 4.2.2. The above isomorphisms are $\mathbf{T}$-*equivariant.* $\mathrm{Hom}(\mathbf{T}, \mathbf{C})$ is a $\mathbf{T}$-module if we let $T \in \mathbf{T}$ act on $\varphi \in \mathrm{Hom}(\mathbf{T}, \mathbf{C})$ by $(T \cdot \varphi)(T') = \varphi(TT')$. If $\alpha : V \to \mathrm{Hom}(\mathbf{T}, \mathbf{C})$ is the above isomorphism (so $\alpha : f \mapsto \varphi_f := (T' \mapsto a_1(f|T'))$) then equivariance is the statement that $\alpha(Tf) = T\alpha(f)$. This follows since

$$\begin{aligned} \alpha(Tf)(T') &= \varphi_{Tf}(T') = a_1(Tf|T') = a_1(f|T'T) \\ &= \varphi_f(T'T) = T\varphi(T') = T\alpha(f)(T'). \end{aligned}$$

4.3 Eigenforms

We continue to assume that $k \geq 2$.

Definition 4.3.1. A modular form $f \in M_k$ is an *eigenform for* $\mathbf{T}$ if $f|T_n = \lambda_n f$ for all $n \geq 1$ and some complex numbers λ_n.

Let f be an eigenform, then $a_n(f) = a_1(f|T_n) = \lambda_n a_1(f)$ so if $a_1(f) = 0$ then $a_n(f) = 0$ for all $n \geq 1$ so since $k \geq 2$ this would imply $f = 0$. Thus $a_1(f) \neq 0$ and we may as well divide through by $a_1(f)$ to obtain the *normalized eigenform* $\frac{1}{a_1(f)} f$. We thus assume that $a_1(f) = 1$, then the formula becomes $a_n(f) = \lambda_n$ and so $f|T_n = a_n(f)f$, for all $n \geq 1$.

Theorem 4.3.2. *Let $f \in V$ and let ψ be the image of f in* $\mathrm{Hom}(\mathbf{T}, \mathbf{C})$*, thus $\psi(T) = a_1(f|T)$. Then f is a normalized eigenform iff ψ is a*
ring homomorphism.

Proof. First suppose f is a normalized eigenform so $f|T_n = a_n(f)f$. Then

$$\begin{aligned}
\psi(T_n T_m) &= a_1(f|T_n T_m) = a_m(f|T_n) \\
&= a_m(a_n(f)f) = a_m(f)a_n(f) \\
&= \psi(T_n)\psi(T_m),
\end{aligned}$$

so ψ is a homomorphism.

Conversely, assume ψ is a homomorphism. Then $f|T_n = \sum a_m(f|T_n)q^m$, so to show that $f|T_n = a_n(f)f$ we must show that $a_m(f|T_n) = a_n(f)a_m(f)$. Recall that $\psi(T_n) = a_1(f|T_n) = a_n$, thus

$$\begin{aligned}
a_n(f)a_m(f) &= a_1(f|T_n)a_1(f|T_m) = \psi(T_n)\psi(T_m) \\
&= \psi(T_n T_m) = a_1(f|T_n|T_m) \\
&= a_m(f|T_n)
\end{aligned}$$

as desired.

□

Chapter 5

Rationality and Integrality Questions

5.1 Review

In the previous lecture we looked at subspaces $V \subset M_k \subset \mathbf{C}[[q]]$, $(k \geq 4)$, and considered the space $\mathbf{T} = \mathbf{T}(V) = \mathbf{C}[\dots, T_n, \dots] \subset \operatorname{End}_{\mathbf{C}} V$ of Hecke operators on V. We defined a pairing $\mathbf{T} \times V \to \mathbf{C}$ by $(T, f) \mapsto a_1(f|T)$ and showed this pairing is nondegenerate and that it induces isomorphisms $\mathbf{T} \cong \operatorname{Hom}(V, \mathbf{C})$ and $V \cong \operatorname{Hom}(\mathbf{T}, \mathbf{C})$.

5.2 Integrality

Fix $k \geq 4$ and let $S = S_k$ be the space of weight k cusp forms with respect to the action of $\mathrm{SL}_2(\mathbf{Z})$. Let

$$S(\mathbf{Q}) = S_k \cap \mathbf{Q}[[q]]$$
$$S(\mathbf{Z}) = S_k \cap \mathbf{Z}[[q]].$$

Theorem 5.2.1. *There is a $\mathbf{C}$-basis of M_k consisting of forms with integral coefficients.*

Proof. This is seen by exhibiting a basis. Recall that for all $k \geq 4$

$$G_k = -\frac{b_k}{2k} + \sum_{k=1}^{\infty} \sum_{d|k} d^{k-1} q^n$$

is the kth Eisenstein series which is a modular form of weight k and

$$E_k = -\frac{2k}{b_k} \cdot G_k = 1 + \cdots$$

is its normalization. Since the Bernoulli numbers $b_2, \dots, b_8$ have 1 as numerator (this isn't always the case, $b_{10} = \frac{5}{66}$) we see that E_4 and E_6 have coefficients in $\mathbf{Z}$ and constant term 1. Furthermore one shows by dimension and independence arguments that the forms

$$\{E_4^a E_6^b | 4a + 6b = k\}$$

form a basis for M_k. $\square$

5.3 Victor Miller's Thesis

Let $d = \dim_{\mathbf{C}} S_k$, then Victor Miller showed in his thesis (see [11], ch. X, theorem 4.4) that there exists

$$f_1, \dots, f_d \in S_k(\mathbf{Z}) \quad \text{such that} \quad a_i(f_j) = \delta_{ij}$$

for $1 \leq i, j \leq d$. The f_i clearly form a basis.

Proposition 5.3.1. *Let $R = \mathbf{Z}[\dots, T_n, \dots] \subset End(S_k)$, then $R = \bigoplus_{i=1}^{d} \mathbf{Z}T_i$.*

Proof. To see that $T_1, \cdots, T_d \in \mathbf{T} = \mathbf{T}(S_k)$ are linearly independent over $\mathbf{C}$ suppose $\sum_{i=1}^{d} c_i T_i = 0$, then

$$0 = a_1(f_j | \sum c_i T_i) = \sum_i c_i a_i(f_j) = \sum_i c_i \delta_{ij} = c_j.$$

From the isomorphism $\mathbf{T} \cong \mathrm{Hom}(S_k, \mathbf{C})$ we know that $\dim_{\mathbf{C}} \mathbf{T} = d$, so we can write any T_n as a $\mathbf{C}$-linear combination

$$T_n = \sum_{i=1}^{d} c_{n_i} T_i, \quad c_{n_i} \in \mathbf{C}.$$

But

$$\mathbf{Z} \ni a_n(f_j) = a_1(f_j | T_n) = \sum_{i=1}^{d} c_{n_i} a_1(f_j | T_i) = \sum_{i=1}^{d} c_{n_i} a_i(f_j) = c_{n_j}$$

so the c_{n_i} all lie in $\mathbf{Z}$ which completes the proof. $\square$

Thus R is an integral Hecke algebra of finite rank d over $\mathbf{Z}$. We have a map

$$\begin{aligned} S(\mathbf{Z}) \times R &\to \mathbf{Z} \\ (f, T) &\mapsto a_1(f|T) \end{aligned}$$

which induces an embedding

$$S(\mathbf{Z}) \hookrightarrow \mathrm{Hom}(R, \mathbf{Z}) \cong \mathbf{Z}^d.$$

Exercise 5.3.2. Prove that the map $S(\mathbf{Z}) \hookrightarrow \mathrm{Hom}(R, \mathbf{Z})$ is in fact an isomorphism of $\mathbf{T}$-modules. [Hint: Show the cokernel is torsion free.]

5.4 Petersson Inner Product

The main theorem is

Theorem 5.4.1. *The $T_n \in \mathbf{T}(S_k)$ are all diagonalizable over $\mathbf{C}$.*

To prove this we note that S_k supports a non-degenerate positive definite Hermitean inner product (the Petersson inner product)

$$(f, g) \mapsto \langle f, g \rangle \in \mathbf{C}$$

such that $\langle f | T_n, g \rangle = \langle f, g | T_n \rangle$. We need some background facts.

Definition 5.4.2. An operator T is *normal* if it commutes with its adjoint, thus $TT^* = T^*T$.

T_n is clearly normal since $T_n^* = T_n$,

Theorem 5.4.3. *A normal operator is diagonalizable.*

Thus each T_n is diagonalizable.

Theorem 5.4.4. *A commuting family of semisimple (=diagonalizable) operators can be simultaneously diagonalized.*

Since the T_n commute this implies S_k has a basis consisting of normalized eigenforms f. Their eigenvalues are real since

$$\begin{aligned} a_n(f)\langle f, f\rangle = \langle a_n(f)f, f\rangle &= \langle f|T_n, f\rangle \\ &= \langle f, a_n(f)f\rangle = \overline{a_n(f)}\langle f, f\rangle. \end{aligned}$$

Exercise 5.4.5. The coefficients a_n of the eigenforms are totally real algebraic integers. [Hint: The space S_k is stable under the action of $\mathrm{Aut}(\mathbf{C})$ on coefficients: if $f = \sum_{n=1}^{\infty} c_n q^n \in S_k$ and $\sigma \in \mathrm{Aut}(\mathbf{C})$ then $\sigma(f) = \sum_{n=1}^{\infty} \sigma(c_n)q^n$ is again in S_k (check this by writing f in terms of a basis $f_1, \ldots, f_d \in S(\mathbf{Z})$). Next use the fact that f is an eigenform iff $\sigma(f)$ is an eigenform.]

Let

$$\mathcal{H} = \{x + iy : x, y \in \mathbf{R}, \text{ and } y > 0\}$$

be the upper half plane. Then the volume form $\frac{dx\wedge dy}{y^2}$ is invariant under the action of

$$\mathrm{GL}_2^+(\mathbf{R}) = \{M \in \mathrm{GL}_2(\mathbf{R}) | \det(M) > 0\}.$$

If $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbf{R})$ then $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ acts on $\mathcal{H}$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : \quad z \mapsto \frac{az+b}{cz+d}$$

and one has

$$\mathrm{Im}\left(\frac{az+b}{cz+d}\right) = \frac{\det(\alpha)}{|cz+d|^2}y.$$

Differentiating $\frac{az+b}{cz+d}$ gives

$$\begin{aligned} d\left(\frac{az+b}{cz+d}\right) &= \frac{a(cz+d)dz - c(az+b)dz}{(cz+d)^2} \\ &= \frac{(ad-bc)dz}{(cz+d)^2} \\ &= \frac{det(\alpha)}{(cz+d)^2}dz \end{aligned}$$

Thus, under the action of α, $dz \wedge d\overline{z}$ takes on a factor of

$$\frac{\det(\alpha)^2}{(cz+d)^2(c\overline{z}+d)^2} = \left(\frac{\det(\alpha)}{|cz+d|^2}\right)^2.$$

Definition 5.4.6. The *Petersson inner product* of forms $f, g \in S_k$ is defined by

$$< f, g >= \int_{\Gamma\backslash\mathcal{H}} (f(z)\overline{g(z)}y^k)\frac{dx \wedge dy}{y^2},$$

where $\Gamma = \mathrm{SL}_2(\mathbf{Z})$.

Integrating over $\Gamma\backslash\mathcal{H}$ can be taken to mean integrating over a fundamental domain for the action of $\mathcal{H}$. Showing that the operators T_n are self-adjoint with respect to the Petersson inner product is a harder computation than Serre [24] might lead one to believe — it takes a bit of thought.

Chapter 6

Modular Curves

6.1 Cusp Forms

Recall that if N is a positive integer we define the congruence subgroups $\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N)$ by

$$\Gamma_0(N) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z}) : c \equiv 0 \pmod{N}\}$$
$$\Gamma_1(N) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z}) : a \equiv d \equiv 1, c \equiv 0 \pmod{N}\}$$
$$\Gamma(N) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z}) : \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \pmod{N}\}.$$

Let Γ be one of the above subgroups. One can give a construction of the space $S_k(\Gamma)$ of cusp forms of weight k for the action of Γ using the language of algebraic geometry. Let $X_\Gamma = \Gamma\backslash\mathcal{H}^*$ be the compactifaction of the upper half plane (union the cusps) modulo the action of Γ. Then X_Γ can be given the structure of Riemann surface and $S_2(\Gamma) = H^0(X_\Gamma, \Omega^1)$ where Ω^1 is the sheaf of differential 1-forms on X_Γ. This works since an element of $H^0(X_\Gamma, \Omega^1)$ is a differential form $f(z)dz$, holomorphic on $\mathcal{H}$ and the cusps, which is invariant with respect to the action of Γ. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ then

$$d(\gamma(z))/dz = (cz+d)^{-2}$$

so

$$f(\gamma(z))d(\gamma(z)) = f(z)dz$$

iff f satisfies the modular condition

$$f(\gamma(z)) = (cz+d)^2 f(z).$$

There is a similar construction of S_k for $k > 2$.

6.2 Modular Curves

One knows that $\mathrm{SL}_2(\mathbf{Z})\backslash\mathcal{H}$ parameterizes isomorphism classes of elliptic curves. The other congruence subgroups also give rise to similar parameterizations. Thus $\Gamma_0(N)\backslash\mathcal{H}$ parameterizes pairs (E, C) where E is an elliptic curve and C is a cyclic subgroup of order N, and $\Gamma_1(N)\backslash\mathcal{H}$ parameterizes pairs (E, P) where E is an elliptic curve and P is a point of exact order N. Note that one can also give a point of exact order N by giving an injection $\mathbf{Z}/N\mathbf{Z} \hookrightarrow E[N]$ or equivalently an injection $\boldsymbol{\mu}_N \hookrightarrow E[N]$ where $\boldsymbol{\mu}_N$ denotes the Nth roots of unity. $\Gamma(N)\backslash\mathcal{H}$ parameterizes pairs $(E, \{\alpha, \beta\})$ where $\{\alpha, \beta\}$ is a basis for $E[N] \cong (\mathbf{Z}/N\mathbf{Z})^2$.

The above quotients spaces are called *moduli spaces* for the *moduli problem* of determining equivalence classes of pairs ($E+$ extra structure).

6.3 Classifying $\Gamma(N)$-structures

Definition 6.3.1. Let S be an arbitrary scheme. An **elliptic curve** E/S is a proper smooth curve

$$\begin{array}{c} E \\ \downarrow f \\ S \end{array}$$

with geometrically connected fibers all of genus one, give with a section "0".

Loosely speaking, proper is a generalization of projective and smooth generalizes non-singularity. See Hartshorne [7], chapter III, section 10, for the precise definitions.

Definition 6.3.2. Let S be any scheme and E/S an elliptic curve. A **$\Gamma(N)$-structure** on E/S is a group homomorphism

$$\varphi : (\mathbf{Z}/N\mathbf{Z})^2 \to E[N](S)$$

whose image "generates" $E[N](S)$.

A good reference is chapter 3 of Katz and Mazur [9].

Define a functor from the category of $\mathbf{Q}$-schemes to the category of sets by sending a scheme S to the set of isomorphism classes of pairs

$$(E, \Gamma(N)\text{-structure})$$

where E is an elliptic curve defined over S and isomorphisms (preserving the $\Gamma(N)$-structure) are taken over S. An isomorphism preserves the $\Gamma(N)$-structure if it takes the two distinguished generators to the two distinguished generators in the image (in the correct order).

Theorem 6.3.3. *For $N \geq 4$ the functor defined above is representable and the object representing it is the modular curve X corresponding to $\Gamma(N)$.*

What this means is that given a $\mathbf{Q}$-scheme S, the set $X(S) = \mathrm{Mor}_{\mathbf{Q}\text{-schemes}}(S, X)$ is isomorphic to the image of the functor's value on S.

There is a natural way to map a pair $(E, \Gamma(N)\text{-structure})$ to an Nth root of unity. If P, Q are the distinguished basis of $E[N]$ we send the pair $(E, \Gamma(N)\text{-structure})$ to

$$e_N(P, Q) \in \boldsymbol{\mu}_N$$

where $e_N : E[N] \times E[N] \to \boldsymbol{\mu}_N$ is the Weil pairing. For the definition of this pairing see chapter III, section 8 of Silverman [33]. The Weil pairing is bilinear, alternating, non-degenerate, Galois invariant, and maps surjectively onto $\boldsymbol{\mu}_N$.

6.4 More on Integral Hecke Operators

We are considering the algebra of integral Hecke operators $\mathbf{T} = \mathbf{T}_{\mathbf{Z}}$ on the space of cusp forms $S_k(\mathbf{C})$ with respect to the action of the full modular group $\mathrm{SL}_2(\mathbf{Z})$. Our goal is to see why $\mathbf{T} \cong \mathbf{Z}^d$ where $d = \dim_{\mathbf{C}} S_k(\mathbf{C})$.

Suppose $A \subset \mathbf{C}$ is any *subring* of $\mathbf{C}$ and recall that

$$\mathbf{T}_A = A[\ldots, T_n, \ldots] \subset \mathrm{End}_{\mathbf{C}} S_k.$$

We have a natural map

$$\mathbf{T}_A \otimes_A \mathbf{C} \to \mathbf{T}_C$$

but we do not yet know that it is an isomorphism.

6.5 Complex Conjugation

We have a conjugate linear map on functions

$$f(\tau) \mapsto \overline{f(-\overline{\tau})}.$$

Since $\overline{(e^{-2\pi i\overline{\tau}})} = e^{2\pi i\tau}$, it follows that

$$\sum_{n=1}^{\infty} a_n q^n \mapsto \sum_{n=1}^{\infty} \overline{a_n} q^n$$

so it is reasonable to call this map "complex conjugation". Furthermore, if we know that

$$S_k(\mathbf{C}) = \mathbf{C} \otimes_{\mathbf{Q}} S_k(\mathbf{Q})$$

then it follows that complex conjugation takes $S_k(\mathbf{C})$ into $S_k(\mathbf{C})$. To see this note that if we have the above equality then every element of $S_k(\mathbf{C})$ is a $\mathbf{C}$-linear combination of elements of $S_k(\mathbf{Q})$ and conversely, and it is clear that the set of such $\mathbf{C}$-linear combinations is invariant under the action of complex conjugation.

6.6 Isomorphism in the Real Case

Proposition 6.6.1. $\mathbf{T}_{\mathbf{R}} \otimes_{\mathbf{R}} \mathbf{C} \cong \mathbf{T}_{\mathbf{C}}$, *as* $\mathbf{C}$*-vector spaces.*

Proof. Since $S_k(\mathbf{R}) = S_k(\mathbf{C}) \cap \mathbf{R}[[q]]$ and since theorem 5.1 assures us that there is a $\mathbf{C}$-basis of $S_k(\mathbf{C})$ consisting of forms with integral coefficients, we see that $S_k(\mathbf{R}) \cong \mathbf{R}^d$ where $d = \dim_{\mathbf{C}} S_k(\mathbf{C})$. (Any element of $S_k(\mathbf{R})$ is a $\mathbf{C}$-linear combination of the integral basis, hence equating real and imaginary parts, an $\mathbf{R}$-linear combination of the integral basis, and the integral basis stays independent over $\mathbf{R}$.) By considering the explicit formula for the action of the Hecke operators T_n on S_k (see section 3) we see that $\mathbf{T}_{\mathbf{R}}$ leaves $S_k(\mathbf{R})$ invariant, thus

$$\mathbf{T}_{\mathbf{R}} = \mathbf{R}[\ldots, T_d, \ldots] \subset \operatorname{End}_{\mathbf{R}} S_k(\mathbf{R}).$$

In section 4 we defined a "perfect" pairing

$$\mathbf{T}_{\mathbf{C}} \times S_k(\mathbf{C}) \to \mathbf{C}$$

which allowed us to show that $\mathbf{T}_{\mathbf{C}} \cong S_k(\mathbf{C})$. By restricting to $\mathbf{R}$ we again get a perfect pairing so we see that $\mathbf{T}_{\mathbf{R}} \cong S_k(\mathbf{R}) \cong \mathbf{R}^d$ which implies that

$$\mathbf{T}_R \otimes_{\mathbf{R}} \mathbf{C} \xrightarrow{\sim} \mathbf{T}_{\mathbf{C}}.$$

$\square$

This also shows that $S_k(\mathbf{C}) \cong \mathbf{C} \otimes_{\mathbf{R}} S_k(\mathbf{R})$ so we have complex conjugation over $\mathbf{R}$.

6.7 The Eichler-Shimura Isomorphism

Our goal in this section is to outline a homological interpretation of S_k. For details see chapter 6 of Lang [11], the original paper of Shimura [32], or chapter VIII of Shimura [31].

How is $S_k(\mathbf{C})$ sort of isomorphic to $H^1(X_\Gamma, \mathbf{R})$? Suppose $k = 2$ and $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ is a congruence subgroup, let $X_\Gamma = \overline{\Gamma\backslash\mathcal{H}}$ be the Riemann surface obtained by compactifying the upper half plane modulo the action of Γ. Then $S_k(\mathbf{C}) = H^0(X_\Gamma, \Omega^1)$ so we have a pairing

$$H_1(X_\Gamma, \mathbf{Z}) \times S_k(\mathbf{C}) \to \mathbf{C}$$

given by integration

$$(\gamma, \omega) \mapsto \int_\gamma \omega.$$

This gives an embedding

$$\mathbf{Z}^{2d} \cong H_1(X_\Gamma, \mathbf{Z}) \hookrightarrow \mathrm{Hom}_{\mathbf{C}}(S_k(\mathbf{C}), \mathbf{C}) \cong \mathbf{C}^d$$

of a "lattice" in $\mathbf{C}^d$. (We say "lattice" since there were some comments by Ribet that Z^{2d} isn't a lattice because the rank might be too small since a subring of $\mathbf{C}^d$ having $\mathbf{Z}$-rank $2d$ might not spans $\mathbf{C}^d$ over $\mathbf{C}$). Passing to the quotient (and compactifying) gives a complex torus called the Jacobian of X_Γ. Again using the above pairing we get an embedding

$$\mathbf{C}^d \cong S_k(\mathbf{C}) \hookrightarrow \mathrm{Hom}(H_1(X_\Gamma, \mathbf{Z}), \mathbf{C}) \cong \mathbf{C}^{2d}$$

which, upon taking the real part, gives

$$S_k(\mathbf{C}) \to \mathrm{Hom}(H_1(X_\Gamma, \mathbf{Z}), \mathbf{R}) \cong H^1(X_\Gamma, \mathbf{R}) \cong H^1_p(\Gamma, \mathbf{R})$$

where $H^1_p(\Gamma, \mathbf{R})$ denotes the *parabolic* group cohomology of Γ with respect to the trivial action. It is this result, that we may view $S_k(\mathbf{C})$ as the cohomology group $H^1_p(\Gamma, \mathbf{R})$, that was alluded to above.

Shimura generalized this for arbitrary $k \geq 2$ so that

$$S_k(\mathbf{C}) \cong H^1_p(\Gamma, V_k)$$

where V_k is a $k-1$ dimensional $\mathbf{R}$-vector space. The isomorphism is (approximately) the following: $f \in S_k(\mathbf{C})$ is sent to the map

$$\gamma \mapsto \mathrm{Re} \int_{\tau_0}^{\gamma\tau_0} f(\tau)\tau^i d\tau, \quad i = 0, \ldots, k-2.$$

Let $W = \mathbf{R} \oplus \mathbf{R}$, then Γ acts on W by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$$

so Γ acts on

$$V_k = \mathrm{Sym}^{k-2} W = W^{\otimes k-2}/S_{k-2}$$

where S_{k-2} is the symmetric group on $k-2$ symbols (note that $\dim V_k = k-1$). Let

$$L = H^1_p(\Gamma, \mathrm{Sym}^{k-2}(\mathbf{Z} \oplus \mathbf{Z}))$$

then under the isomorphism

$$S_k(\mathbf{C}) \cong H^1_p(\Gamma, \mathbf{R})$$

L is a sublattice of $S_k(\mathbf{C})$ of $\mathbf{Z}$-rank 2 which is T_n-stable for all n. Thus we have an embedding

$$\mathbf{T}_{\mathbf{Z}} = \mathbf{T} \hookrightarrow \mathrm{End}\, L$$

and so $\mathbf{T}_{\mathbf{R}} \subset \mathrm{End}_{\mathbf{R}}(L \otimes \mathbf{R})$ and $\mathbf{T}_{\mathbf{Z}} \otimes_{\mathbf{Z}} \mathbf{R} \cong \mathbf{T}_{\mathbf{R}}$ which has rank d.

6.8 The Petterson Inner Product is Hecke Compatible

Theorem 6.8.1. *Let* $\Gamma = \mathrm{SL}_2(\mathbf{Z})$*, let* $f, g \in S_k(\mathbf{C})$*, and let*

$$\langle f, g\rangle = \int_{\Gamma\backslash\mathcal{H}} f(\tau)\overline{g(\tau)}y^k \frac{dxdy}{y^2}.$$

Then this integral is well-defined and Hecke compatible, that is, $\langle f|T_n, g\rangle = \langle f, g|T_n\rangle$ *for all* n*.*

Proof. See Chapter 3 of Lang [11]. □

Chapter 7

Higher Weight Modular Forms

We are considering the spaces $S_k(\mathbf{C})$, $S_k(\mathbf{Q})$ and $S_k(\mathbf{Z})$ which all have rank d. Each is acted upon by the Hecke algebra $\mathbf{T}$. We defined a Hecke compatible inner product (the Peterrson product) and used it to show that

$$S_k(\mathbf{Z}) \cong \mathrm{Hom}_{\mathbf{Z}}(\mathbf{T}, \mathbf{Z}).$$

7.1 Definitions of T

> "I may be asking you to explain something we have already discussed, but have we intrinsically defined the Hecke operators yet?"
>
> – Saul Schliemer

The T_n are defined as operators on $S_k(\mathbf{C})$ by defining their action on modular forms and noting from explicit formulas that $S_k(\mathbf{C})$ is preserved. But the T_n can be thought of in other ways, for example, since

$$S_k(\mathbf{C}) \cong H^1_p(\Gamma, \mathrm{Sym}^{k-2}(\mathbf{R} \oplus \mathbf{R}))$$

one may give an *explicit* description of the action of T_n on H^1_p.

7.2 Double Cosets

Let p be a prime and

$$\mathcal{M}_p = \{\alpha \in M_2(\mathbf{Z}) : \det(\alpha) = p\}.$$

Let $F : \mathcal{L} \to \mathbf{C}$ be a function on the free abelian group of lattices and recall that T_p acts on F by

$$(T_p F)(L) = p^{k-1} \sum_{\substack{L' \subset L \\ (L:L')=p}} F(L').$$

One can write $\mathcal{M}_p$ as a disjoint union of left cosets,

$$\mathcal{M}_p = \mathrm{SL}_2(\mathbf{Z}) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \mathrm{SL}_2(\mathbf{Z}) = \bigcup_{\substack{ad=p \\ 0 \le b < d}} \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mathrm{SL}_2(\mathbf{Z}) \,.$$

Then $\sum_{(L:L')=p} L'$ may be thought of as the sum of the images of L under the action of the left cosets of $\mathcal{M}_p$. For a complete exposition, in greater generality, see Shimura [31], especially chapter 3.

7.3 More General Congruence Subgroups

Definition 7.3.1. A **Dirichlet character** mod N is a homomorphism

$$\varepsilon : (\mathbf{Z}/n\mathbf{Z})^* \to \mathbf{C}^*$$

extended to $\mathbf{Z}/N\mathbf{Z}$ by putting $\varepsilon(m) = 0$ if $(m, N) \neq 1$.

Fix integers $k \geq 0$ and $N \geq 1$. In this section we consider the spaces

$$S_k(\Gamma_1(N), \epsilon)$$

for Dirichlet characters ε mod N and explicitly describe the action of the Hecke operators

$$\begin{cases} T_n, & n \geq 1 \\ \langle d \rangle, & d \in (\mathbf{Z}/N\mathbf{Z})^* \end{cases}$$

on these spaces.

Remark 7.3.2. Let n be a positive integer. If $(n, N) = 1$, then the T_n behave like they do for $\mathrm{SL}_2(\mathbf{Z})$. In fact, the T_n and $\langle d \rangle$ commute and

$$(f|T_n, g) = (f, g|\langle n \rangle^{-1} T_n)$$

$$(f|\langle d \rangle, g) = (f, g|\langle d \rangle^{-1})$$

so the T_n (for n prime to N) and $\langle d \rangle$ are simultaneously diagonalizable. But if $(n, N) \neq 1$ then T_n may not be diagonalizable.

Definition 7.3.3. Let

$$S_k(\Gamma_1(N)) = \{f : f(\gamma\tau) = (c\tau + d)^k f(\tau) \text{ all } \gamma \in \Gamma_1(N)\}$$

where the f are assumed holomorphic on $\mathcal{H} \cup \{\text{cusps}\}$. For each Dirichlet character ε mod N let

$$S_k(\Gamma_1(N), \varepsilon) = \{f : f(\gamma\tau)(c\tau + d)^{-k} = \varepsilon(d) f \text{ all } \gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)\}.$$

When $\varepsilon \neq 0$ and $f \in S_k(\Gamma_1(N), \varepsilon)$ one calls ε the **nebentypus** of f.

Let $d \in (\mathbf{Z}/N\mathbf{Z})^*$ and let $f \in S_k(\Gamma_1(N))$. Let $\gamma = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \in \Gamma_1(N)$ be a matrix whose lower right entry is congruent to d mod N. Then we define

$$f(\tau)|\langle d \rangle = f(\gamma\tau)(c\tau + d)^{-k}.$$

Since $f|\langle d \rangle = \varepsilon(d) f$, $S_k(\Gamma_1(N), \varepsilon)$ is the $\varepsilon(d)$ eigenspace of $\langle d \rangle$ and $\langle d \rangle$ is diagonalizable so one has a direct sum decomposition

$$S_k(\Gamma_1(N)) = \bigoplus_{\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*} S_k(\Gamma_1(N), \varepsilon).$$

If $f \in S_k(\Gamma_1(N), \varepsilon)$ then

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \in \Gamma_0(N)$$

so

$$f(-\tau)(-1)^{-k} = \varepsilon(-1) f(\tau)$$

so that $S_k(\Gamma_1(N), \varepsilon) = 0$ unless $\varepsilon(-1) = (-1)^k$. Thus about half of the direct summands vanish.

7.4 Explicit Formulas

Let

$$f = \sum_{n=1}^{\infty} a_n q^n \in S_k(\Gamma_1(N), \varepsilon)$$

and let p be a prime, then

$$f|T_p = \begin{cases} \sum_{n=1}^{\infty} a_{np} q^n + p^{k-1}\varepsilon(p) \sum_{n=1}^{\infty} a_n q^{pn}, & p \nmid N \\ \sum_{n=1}^{\infty} a_{np} q^n + 0, & p | N \end{cases}$$

When $p|N$, T_p is often denoted U_p and called an Atkin-Lehner operator.

We have the relations

$$T_m T_n = T_{mn}, \quad (m, n) = 1$$
$$T_{p^k} = \begin{cases} (T_p)^k, & p|N \\ ?, & p \nmid N \end{cases}$$

7.5 Old and New Forms

Warning: T_p is not necessarily diagonalizable if $p|N$. There is an example due to Shimura, to present it we must first introduce old and new forms.

Let M and N be positive integers such that $M|N$ and let $d|\frac{N}{M}$. If $f(\tau) \in S_k(\Gamma_1(M))$ then $f(d\tau) \in S_k(\Gamma_1(N))$. We thus have a map $S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N))$ for each $d|\frac{N}{M}$. Combining these gives a map

$$\varphi_M : \bigoplus_{d|\frac{N}{M}} S_k(\Gamma_1(d)) \to S_k(\Gamma_1(N)).$$

Definition 7.5.1. The **old part** of $S_k(\Gamma_1(N))$ is the subspace generated by the images of the φ_M for $M|N$, $M \neq N$.

We remark that the **new part** of $S_k(\Gamma_1(N))$ is the orthogonal complement of the old part with respect to the Petersson inner product.

Chapter 8

New Forms

Today we discuss how the Hecke operators T_n on $S_k(\Gamma_1(N))$ can fail to be diagonalizable. Let N be a positive integer and M a divisor of N. For each $d|\frac{N}{M}$ we define a map

$$\alpha_d : S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N)) : \quad f(\tau) \mapsto f(d\tau).$$

Note that when T_p acts on the image space $S_k(\Gamma_1(N))$ we will often denote it by U_p. We must check that $f(d\tau) \in S_k(\Gamma_1(N))$. Define for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$,

$$(f|[\gamma]_k)(\tau) = \det(\gamma)^{k-1}(cz+d)^{-k} f(\gamma(\tau)).$$

Thus $f \in S_k(\Gamma_1(N))$ iff $f|[\gamma]_k(\tau) = f(\tau)$ (and f is holomorphic). Now let $f(\tau) \in \Gamma_1(M)$ and let $\iota_d = \left(\begin{smallmatrix} d & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then $f|[\iota_d]_k(\tau) = d^{k-1} f(d\tau)$ is a modular form on $\Gamma_1(N)$ since $\iota_d^{-1}\Gamma_1(M)\iota_d$ contains $\Gamma_1(N)$ (check this directly by conjugating an element of $\Gamma_1(N)$ by ι_d). Moreover if f is a cusp form then so is $f|[\iota_d]_k$. If $f \in S_k(\Gamma_1(M))$ is nonzero, then as d varies over divisors of $\frac{N}{M}$, the various $f(d\tau)$ are linearly independent.

Suppose $f \in S_k(\Gamma_1(M))$ is a normalized eigenform for all of the Hecke operators T_n and $\langle n \rangle$, and p is a prime not dividiing M. Then

$$f|T_p = af \quad \text{and} \quad f|\langle p \rangle = \varepsilon(p) f.$$

Assume $N = p^r M$ where r is an integer ≥ 1. Let

$$f_i(\tau) = f(p^i \tau),$$

so $f_0, \dots, f_r$ are the images of f under the maps defined above and $f = f_0$. Consider the action of U_p on the f_i. From previous work we have

$$\begin{aligned} f|T_p &= \sum_{n \geq 1} a_{np} q^n + \varepsilon(p) p^{k-1} \sum a_n q^{pn} \\ &= f_0|U_p + \varepsilon(p) p^{k-1} f_1 \end{aligned}$$

so

$$f_0|U_p = f|T_p - \varepsilon(p) p^{k-} f_1 = af_0 - \varepsilon(p) p^{k-1} f_1.$$

Also

$$f_1|U_p = (\sum a_n q^{pn})|U_p = \sum a_n q^n = f_0.$$

More generally one can show that $f_i|U_p = f_{i-1}$.

U_p preserves the two dimensional vector space spanned by f_0 and f_1. The matrix of U_p is

$$A = \begin{pmatrix} a & 1 \\ -\varepsilon(p)p^{k-1} & 0 \end{pmatrix}$$

which has characteristic polynomial

$$\chi_A(X) = X^2 - aX + p^{k-1}\varepsilon(p).$$

8.1 Connection With Galois Representations

This leads to a striking connection with Galois representations. Let f be a modular form and E be the field generated over $\mathbf{Q}$ by the coefficients of f. Let ℓ be a prime and λ a prime lying over ℓ. Then one constructs a representation

$$\rho_\lambda : \mathcal{G}al(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, E_\lambda).$$

If $p \not| N\ell$, then ρ_λ is unramified at p, so there is a Frobenious element $\mathrm{frob}_p \in \mathcal{G}al(\overline{\mathbf{Q}}/\mathbf{Q})$. One can show that

$$\det(\rho_\lambda(\mathrm{frob}_p)) = p^{k-1}\varepsilon(p)$$
$$\mathrm{Tr}(\rho_\lambda(\mathrm{frob}_p)) = a_p = a,$$

so the characteristic polynomial of $\rho_\lambda(\mathrm{frob}_p) \in \mathrm{GL}_2(E_\lambda)$ is

$$X^2 - a_pX + p^{k-1}\varepsilon(p).$$

8.2 Semisimplicity of U_p

Question 8.2.1. *Is U_p semisimple on the span of f_0 and f_1?*

If the eigenvalues are distinct the answer is clearly yes. If the eigenvalues are the same, then $X^2 - aX + p^{k-1}\varepsilon(p)$ has discriminant zero, that is, $4\varepsilon(p)p^{k-1} = a^2$ so

$$a = 2p^{\frac{k-1}{2}}\sqrt{\varepsilon(p)}.$$

Is this possible? The answer is still *unknown*, although it is a curious fact that the Ramanujan conjectures (proved by Delign in 1973) imply that $|a| \le 2p^{\frac{k-1}{2}}$, so the above equality remains taunting.

When $k = 2$ Weil showed that $\rho_\lambda(\mathrm{frob}_p)$ is semisimple so if the eigenvalues of U_p are equal then $\rho_\lambda(\mathrm{frob}_p)$ is a scalar. But Edixhoven and Coleman [5] show that it is not a scalar by looking at the abelian variety attached to f.

8.3 Shimura's Example of Nonsemisimple U_p

Let W be the space spanned by f_0, f_1 and let V be the space spanned by f_0, f_1, f_2, f_3. U_p acts on V/W by $\overline{f_2} \mapsto 0$ and $\overline{f_3} \mapsto \overline{f_2}$. Thus the matrix of the action of U_p on V/W is $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ which is nonzero and nilpotent hence not semisimple. Since W is invariant under U_p this shows that U_p is not semisimple on V.

8.4 An Interesting Duality

Now suppose $N = 1$ thus $\Gamma_1(N) = \mathrm{SL}_2(\mathbf{Z})$. Because of the Petersson product all the T_n are diagonalizable, so $S_k = S_k(\Gamma_1(1))$ has a basis

$$f_1, \ldots, f_d$$

of normalized eigenforms where $d = \dim S_k$. Let $\mathbf{T} = \mathbf{T_C}$, then there is a *canonical* map

$$\mathbf{T_C} \hookrightarrow \mathbf{C}^d : \quad T \mapsto (\lambda_1, \ldots, \lambda_d)$$

where $f_i|T = \lambda_i f_i$. This map is clearly injective and we know by previous arguments that $\dim \mathbf{T_C} = d$ so the map is an isomorphism of $\mathbf{C}$-vector spaces.

The form

$$v = f_1 + \cdots + f_n$$

generates S_k as a $\mathbf{T}$-module. Since v corresponds to the vector $(1, \ldots, 1)$ and $\mathbf{T} \cong \mathbf{C}^d$ acts on $S_k \cong \mathbf{C}^d$ componentwise this is just the statement that $\mathbf{C}^d$ is generated by $(1, \ldots, 1)$ as a $\mathbf{C}^d$-module – which is clear. Thus we have simultaneously:

1) S_k is free of rank 1 over $\mathbf{T}$, and
2) $S_k = \mathrm{Hom}_{\mathbf{C}}(\mathbf{T}, \mathbf{C})$ as $\mathbf{T}$-modules, thus

$$\mathbf{T} \cong \mathrm{Hom}_{\mathbf{C}}(\mathbf{T}, \mathbf{C}).$$

The isomorphism sends an element of $T \in \mathbf{T}$ to $Tv \in S_k$. Since the identification $S_k = \mathrm{Hom}_{\mathbf{C}}(\mathbf{T}, \mathbf{C})$ was constructed using the Petersson product it is canonical and since the choice of a normalized eigenbasis $f_1, \ldots, f_d$ is canonical we see that the isomorphism $T \cong \mathrm{Hom}_{\mathbf{C}}(\mathbf{T}, \mathbf{C})$ is canonical.

Proposition 8.4.1. $v \in S_k(\mathbf{Q})$

Proof. Let $\sigma \in \mathcal{G}al(\overline{\mathbf{Q}}/\mathbf{Q})$, then if f_i is a normalized eigenform so is $\sigma(f_i)$ (from the explicit formula). Thus $\sigma(f_1 + \cdots + f_n) = f_1 + \cdots + f_n$ for all σ as desired. $\square$

Now we consider the case for general N. Recall that we have defined maps

$$S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N))$$

for all M dividing N and all divisors d of $\frac{N}{M}$.

Definition 8.4.2. The **old part** of $S_k(\Gamma_1(N))$ is the space generated by all images of these maps with $M|N$ but $M \neq N$. The **new part** is the orthogonal complement of the old part with respect to the Petersson product.

There is an algebraic definition of the new part. One defines certain trace maps

$$S_k(\Gamma_1(N)) \to S_k(\Gamma_1(M))$$

for all $M < N$, $M|N$ which are the adjoints to the above maps (w.r.t Petersson product). Then f is in the new part of $S_k(\Gamma_1(N))$ iff f is killed by all of these maps.

It can be shown that the T_n act semisimply on $S_k(\Gamma_1(M))_{\mathrm{new}}$ for all $M \geq 1$. Thus $S_k(\Gamma_1(M))_{\mathrm{new}}$ has a basis of eigenforms. We have a natural map

$$\bigoplus_{M|N} S_k(\Gamma_1(M))_{\mathrm{new}} \hookrightarrow S_k(\Gamma_1(N)).$$

The image in $S_k(\Gamma_1(N))$ of an eigenform f for some $S_k(\Gamma_1(M))_{\text{new}}$ is called a **newform** of level $M_f = M$. Note that a newform is not necessarily an eigenform for the Hecke operators acting on $S_k(\Gamma_1(N))$. Let

$$v = \sum_f f(q^{\frac{N}{M_f}}) \in S_k(\Gamma_1(N))$$

where the sum is taken over all newforms f of weight k and some level $M|N$. This generalizes the v constructed above when $N = 1$ and has many of the same good properties. For example, $S_k(\Gamma_1(N))$ is free of rank 1 over $\mathbf{T}$ with basis element v. The coefficients of v lie in $\mathbf{Q}$, but to show this we need to know the new part of $S_k(\Gamma_1(N))$ is stable under the action of the Galois group of $\mathbf{Q}$. This is not easy since the new part is defined in terms of the Petersson product which is an analytic construction. Serre circumvents this problem by giving an alternative definition in terms of trace maps going the other way.

8.5 Observations on T_n

Let $\mathbf{T}_{\mathbf{Q}} = \mathbf{Q}[\cdots, T_n, \cdots]$ and $\Gamma = \Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$. Let $f_1, \ldots, f_d$ be a basis of Γ consisting of normalized eigenforms.

Proposition 8.5.1. *The coefficients of the f_i are totally real algebraic integers.*

Proof. $\mathcal{G}al(\mathbf{C}/\mathbf{Q})$ acts on f_i by acting on the coefficients of its q-expansion. From the explicit formula in section 3.2 one sees that the set $\{f_1, \ldots, f_d\}$ is stable under the action of $\mathcal{G}al(\mathbf{C}/\mathbf{Q})$. For any i, $a_n(f_i)$ is an eigenvalue of T_n since $f_i|T_n = a_n(f_i)f_i$, and T_n is self-adjoint so $a_n(f_i)$ must be real. Thus all conjugates of $a_n(f_i)$ are real and there are only finitely many since a conjugate of $a_n(f_i)$ must be $a_n(f_j)$ for some j, $1 \le j \le d$. $\square$

Proposition 8.5.2. *The operators $\langle d\rangle$ on $S_k(\Gamma_1(N))$ lie in $\mathbf{Z}[\ldots, T_n, \ldots]$.*

Proof. It is enough to show $\langle p\rangle \in \mathbf{Z}[\ldots, T_n, \ldots]$ for There is a formula relating $\langle p\rangle$ and T_p,

$$p^{k-1}\langle p\rangle = T_p{}^2 - T_{p^2}.$$

By Dirichlet's theorem on prime's in arithmetic progression, see VIII.4 of Lang [10], there is another prime q congruent to p mod N. Since p^{k-1} and q^{k-1} are relatively prime there exist integers a and b so that $ap^{k-1} + bq^{k-1} = 1$. Then

$$\langle p\rangle = \langle p\rangle(ap^{k-1} + bq^{k-1}) = a(T_p{}^2 - T_{p^2}) + b(T_q{}^2 - T_{q^2}).$$

$\square$

Let Σ be a set of representatives of $\{f_1, \ldots, f_d\}\backslash\mathcal{G}al(\mathbf{C}/\mathbf{Q})$. It is unknown whether or not $\#\Sigma$ can be larger than one, that is, whether the eigenforms are all conjugate under the action of Galois. Let $K_f = \mathbf{Q}(\ldots, a_n(f), \ldots)$ and defined a homomorphism of $\mathbf{Q}$-algebras

$$T_{\mathbf{Q}} \to K_f : T_n \mapsto \lambda \text{ where } T_n f = \lambda f$$

Taking the product over a set of representatives of the f_i yields a map

$$\mathbf{T}_{\mathbf{Q}} \xrightarrow{\sim} \prod_{f\in\Sigma} K_f$$

which one can show is an isomorphism of $\mathbf{Q}$-algebras.

Example 8.5.3. Consider $S_2(\Gamma_0(N))$ with N prime, then

$$\mathbf{T}_{\mathbf{Q}} \cong E_1 \times \cdots E_t$$

with the E_i totally real fields. When $N = 37$, that $\mathbf{T}_{\mathbf{Q}} \cong \mathbf{Q} \times \mathbf{Q}$.

Chapter 9

Some Explicit Genus Computations

9.1 Computing the Dimension of $S_k(\Gamma)$

Let $k = 2$ unless otherwise noted, and let $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ be a congruence subgroup. Then

$$S_2(\Gamma) = H^0(X_\Gamma, \Omega^1)$$

where

$$X_\Gamma = (\Gamma\backslash\mathcal{H}) \cup (\Gamma\backslash\mathbf{P}^1(\mathbf{Q})).$$

By definition $\dim H^0(X_\Gamma, \Omega^1)$ is the genus of X_Γ.

Exercise 9.1.1. Prove that when $\Gamma = \mathrm{SL}_2(\mathbf{Z})$ then $\Gamma\backslash\mathbf{P}^1(\mathbf{Q})$ is a point.

Since $\Gamma \subset \Gamma(1)$ there is a covering

$$\begin{array}{ccccc} \Gamma\backslash\mathcal{H} & \longrightarrow & X_\Gamma & & \\ \downarrow & & \downarrow & & \\ \Gamma(1)\backslash\mathcal{H} & \longrightarrow & X_{\Gamma(1)} & \xrightarrow{j} & \mathbf{P}^1(\mathbf{C}) \end{array}$$

which is only ramified at points above $0, 1728, \infty$ (0 corresponds to i and ρ to 1728 under j).

Example 9.1.2. Suppose $\Gamma = \Gamma_0(N)$, then the degree of the covering is the index $(\mathrm{SL}_2(\mathbf{Z})/\{\pm 1\} : \Gamma_0(N)/\{\pm 1\})$. A point on $Y_{\Gamma(1)}$ corresponds to an elliptic curve, whereas a points on $Y_0(N)$ correspond to a pair consisting of an elliptic curve and a subgroup of order N.

9.2 Application of Riemann-Hurwitz

Now we compute the genus of X_Γ by applying the Riemann-Hurwitz formula. Intuitively the Euler charcteristic should be totally additive, that is, if A and B are disjoint spaces then

$$\chi(A \cup B) = \chi(A) + \chi(B).$$

Let X be a compact Riemann surface of genus g, then $\chi(X) = 2 - 2g$. Since $\chi(\{\text{point}\}) = 1$ we should have that

$$\chi(X - \{p_1, \ldots, p_n\}) = \chi(X) - n\chi(1) = (2 - 2g) - n.$$

If we have an umramified covering $X \to Y$ of degree d then $\chi(X) = d \cdot \chi(Y)$. Consider the covering

$$\begin{array}{c} X_\Gamma - \{\text{points over } 0, 1728, \infty\} \\ \Big\downarrow \\ X_{\Gamma(1)} - \{0, 1728, \infty\} \end{array}$$

Since $X_{\Gamma(1)}$ has genus 0, $X_{\Gamma(1)} - \{0, 1728, \infty\}$ has Euler characteristic $2 - 3 = -1$. If we let $g = \chi(X_\Gamma)$ then $\chi(X_\Gamma - \{\text{points over } 0, 1728, \infty\} = 2 - 2g - n_0 - n_{1728} - n_\infty$, where n_p denotes the number of points lying over p. Thus $-d = 2 - 2g - n_0 - n_{1728} - n_\infty$ whence

$$2g - 2 = d - n_0 - n_{1728} - n_\infty.$$

Suppose $\Gamma = \Gamma_0(N)$ with $N > 3$, then $n_0 = d/3$ and $n_{1728} = d/2$ (I'm not sure why). The degree d of the covering is equal to the number of unordered ordered basis of $E[N]$, thus

$$d = \#\,\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/2.$$

We still need to compute n_∞. $\mathrm{SL}_2(\mathbf{Z})$ acts on $\mathbf{P}^1(\mathbf{Q})$ if we view $\mathbf{P}^1(\mathbf{Q})$ as all pairs (a, b) of relatively prime integers and suppose ∞ corresponds to $(1, 0)$. The stabilizer of $(1, 0)$ is the sugroup $\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z}) : c = 0\}$ of upper triangular matrices. Since the points lying over ∞ are all conjugate by the Galois group of the covering (which is $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$),

$$\text{number of cusps} = \frac{\text{order of } \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}}{\text{order of stabilizer of } \infty}.$$

We thus have

$$2g(X(N)) - 2 = \frac{d}{6} - \frac{d}{N}$$

where $\frac{d}{N}$ is the number of cusps.

9.3 Explicit Genus Computations

Let $N > 3$ and consider the modular curve $X = X(N)$. There is a natural covering map $X \to X(1) \xrightarrow{j} \mathbf{C}$. Let d be the degree, then

$$2g - 2 = d - m_0 - m_{1728} - m_\infty$$

where g is the genus of X and m_x is the number of points lying over x. Since m_0 is approximately $\frac{d}{3}$ and m_{1728} is approximately $\frac{d}{2}$,

$$2g - 2 = \frac{d}{6} - m_\infty \pm \text{ small correction factor.}$$

9.4 The Genus of $X(N)$

Now we count the number of cusps of $X(N)$, that is, the size of $\Gamma(N)\backslash\mathbf{P}^1(\mathbf{Q})$. There is a surjective map from $\mathrm{SL}_2(\mathbf{Z})$ to $\mathbf{P}^1(\mathbf{Q})$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Let U be the kernel, thus U is the stabilizer of $\infty = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$, so $U = \{\pm\left(\begin{smallmatrix}1&a\\0&1\end{smallmatrix}\right) : a \in \mathbf{Z}\}$. Then the cusps of $X(N)$ are the elements of

$$\Gamma(N)\backslash(\mathrm{SL}_2(\mathbf{Z})\,/U) = (\Gamma(N)\backslash\,\mathrm{SL}_2(\mathbf{Z}))/U = \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/U$$

which has order

$$\frac{\#\,\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})}{2N} = \frac{d}{N}.$$

Substituting this into the above formula gives

$$2g-2 = \frac{d}{6} - \frac{d}{N} = \frac{d}{6N}(N-6)$$

so

$$g = 1 + \frac{d}{12N}(N-6).$$

When N is prime

$$d = \frac{1}{2}\#\,\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z}) = \frac{1}{2}\cdot\frac{(N^2-1)(N^2-N)}{N-1}.$$

Thus when $N=5$, $d=60$ so $g=0$, and when $N=7$, $d=168$ so $g=3$.

9.5 The Genus of $X_0(N)$

Suppose $N>3$ and N is prime. The covering map $X_0(N)\to X(1)$ is of degree $N+1$ since a point of $X_0(N)$ corresponds to an elliptic curve along with a subgroup of order N and there are $N+1$ such subgroups because N is prime.

Exercise 9.5.1. $X_0(N)$ has two cusps; they are the orbit of ∞ which is unramified and 0 which is ramified of order N.

Thus

$$2g-2 = N+1-2-n_{1728}-n_0.$$

n_0 is the number of pairs (E,C) (modulo isomorphism) such that E has j-invariant 0. So we consider $E = \mathbf{C}/\mathbf{Z}[\frac{-1+i\sqrt{3}}{2}]$ which has endomorphism ring $\mathrm{End}(E) = \mathbf{Z}[\boldsymbol{\mu}_6]$. Now $\boldsymbol{\mu}_6/\pm 1$ acts on the cyclic subgroups C so, letting ω be a primitive cube root of unity, we have

$$(E,C)\cong(E,\omega C)\cong(E,\omega^2 C).$$

This might lead one to think that m_0 is $(N+1)/3$, but it may be bigger if, for example, $C=\omega C$. Thus we must count those C so that $\omega C = C$ or $\omega^2 C = C$, that is, those C which are stable under $\mathcal{O} = \mathbf{Z}[\frac{-1+i\sqrt{3}}{2}]$. So we must compute the number of stable $\mathcal{O}/N\mathcal{O}$-submodules of order N. This depends on the structure of $\mathcal{O}/N\mathcal{O}$:

$$\mathcal{O}/N\mathcal{O} = \begin{cases}\mathbf{F}_N\oplus\mathbf{F}_N & \text{if } (\frac{-3}{N}) = 1 \text{ (N splits)}\\ \mathbf{F}_{N^2} & \text{if } (\frac{-3}{N}) = -1 \text{ (N stays inert)}\end{cases}$$

Since $\mathcal{O}/N\mathcal{O} = \mathbf{F}_{N^2}$ is a field it has no submodules of order N, whereas $\mathbf{F}_N\oplus\mathbf{F}_N$ has two $\mathcal{O}/N\mathcal{O}$-submodules of order N, namely $\mathbf{F}_N\oplus 0$ and $0\oplus\mathbf{F}_N$. Thus

$$m_0 = \begin{cases}\frac{N+1}{3} & \text{if } N\equiv 2 \pmod 3\\ \frac{N-1}{3}+2 & \text{if } N\equiv 1 \pmod 3\end{cases}$$

Exercise 9.5.2. It is an exercise in elegance to write this as a single formula involving the quadratic symbol.

By similar reasoning one shows that

$$m_{1728} = \begin{cases} \frac{N+1}{2} & \text{if } N \equiv 3 \pmod 4 \\ \frac{N-1}{2} + 2 & \text{if } N \equiv 1 \pmod 4 \end{cases}$$

We can now compute the genus of $X_0(N)$ for any prime N. For example, if $N = 37$ then $2g - 2 = 36 - (2 + 18) - (14) = 2$ so $g = 2$. Similarly, $X_0(13)$ has genus 0 and $X_0(11)$ has genus 1. In general, $X_0(N)$ has genus approximately $N/12$.

Serre constructed a nice formula for the above genus. Suppose $N > 3$ is a prime and write $N = 12a + b$ with $0 \leq b \leq 11$. Then Serre's formula is

b	1	5	7	11
g	$a-1$	a	a	$a+1$

9.6 Modular Forms mod p

Let N be a positive integer, let p be a prime and assume Γ is either $\Gamma_0(N)$ or $\Gamma_1(N)$.

Definition 9.6.1. Let $M_k(\Gamma, \mathbf{Z}) = M_k(\Gamma, \mathbf{C}) \cap \mathbf{Z}[[q]]$, then

$$M_k(\Gamma, \mathbf{F}_p) = M_k(\Gamma, \mathbf{Z}) \otimes_{\mathbf{Z}} \mathbf{F}_p$$

is the space of **modular forms mod** p of weight k.

Suppose $p = N$, then one has **Serre's Equality**:

$$M_{p+1}(\mathrm{SL}_2(\mathbf{Z}), \mathbf{F}_p) = M_2(\Gamma_0(p), \mathbf{F}_p)$$

The map from the right hand side to the left hand side is accomplished via a certain normalized Eisenstein series. Recall that in $\mathrm{SL}_2(\mathbf{Z})$,

$$G_k = -\frac{B_k}{2k} + \sum_{n=1}^{\infty} (\sum_{d|n} d^{k-1}) q^n$$

and

$$E_k = 1 - \frac{2k}{B_k} \sum_{n=1}^{\infty} (\sum_{d|n} d^{k-1}) q^n.$$

One finds $\mathrm{ord}_p(-\frac{B_k}{2k})$ using Kummer congruences. In particular, $\mathrm{ord}_p(B_{p-1}) = -1$, so $E_{p-1} \equiv 1 \pmod p$. Thus multiplication by E_{p-1} raises the level by $p - 1$ but does not change the q-expansion mod p. We thus get a map

$$M_2(\Gamma_0(p), \mathbf{F}_p) \to M_{p+1}(\Gamma_0(p), \mathbf{F}_p).$$

The map

$$M_{p+1}(\Gamma_0(p), \mathbf{F}_p) \to M_{p+1}(\mathrm{SL}_2(\mathbf{Z}), \mathbf{F}_p)$$

is the trace map (which is dual to the natural inclusion going the other way) and is accomplished by averaging in order to get a form invariant under $\mathrm{SL}_2(\mathbf{Z})$.

Chapter 10

The Field of Moduli

In this chapter we will study the field of definition of the modular curves $X(N)$, $X_0(N)$, and $X_1(N)$.

The function field of $X(1) = \mathbf{P}^1_{\mathbf{Q}}$ is $\mathbf{Q}(t)$. If E is an elliptic curve given by a Weierstrass equation $y^2 = 4x^3 - g_2x - g_3$ then

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

The j invariant determines the isomorphism class of E over $\mathbf{C}$. Pick an elliptic curve $E/\mathbf{Q}(t)$ such that $j(E) = t$. In particular we could pick the elliptic curve with Weierstrass equation

$$y^2 = 4x^3 - \frac{27t}{t-1728}x - \frac{27t}{t-1728}.$$

Let E/k be an arbitrary elliptic curve and N a positive integer prime to $\operatorname{char} k$. Then $E[N](\overline{k}) \cong (\mathbf{Z}/N\mathbf{Z})^2$. Let $k(E[N])$ be the field obtained by adjoining the coordinates of the N-torsion points of E. Consider the tower of fields $\overline{k} \supset k(E[N]) \supset k$. There is a Galois representation on the N torsion of E:

$$\mathcal{G}al(\overline{k}/k) \xrightarrow{\rho_{E,N}} \operatorname{Aut}(E[N]) \cong \operatorname{GL}_2(\mathbf{Z}/N\mathbf{Z})$$

and $\mathcal{G}al(\overline{k}/k(E[N])) = \ker(\rho_{E,N})$. Thus the Galois group of the extension $\mathbf{Q}(t)(E[N])$ over $\mathbf{Q}(t)$ is contained in $\operatorname{GL}_2(\mathbf{Z}/N\mathbf{Z})$. Let $X(N)$ be the curve corresponding to the function field $\mathbf{Q}(t)(E[N])$ over $\mathbf{Q}$. Since $\overline{\mathbf{Q}} \cap \mathbf{Q}(t)(E[N])$ is contained in $Q(\boldsymbol{\mu}_N)$, $X(N)$ is defined over $\mathbf{Q}(\boldsymbol{\mu}_N)$.

Composing ρ_E with the natural map $\operatorname{GL}_2(\mathbf{Z}/N\mathbf{Z}) \to \operatorname{GL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$ gives a map

$$\overline{\rho}_E : \mathcal{G}al(\overline{K}/K) \to \operatorname{GL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}.$$

Proposition 10.0.2. *$\overline{\rho}_E$ is surjective iff ρ_E is surjective.*

Proof. If $\overline{\rho}_E$ is surjective then either $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ or its negative lies in the image of ρ. Thus $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ lie in the image of ρ. Since $\overline{\rho}_E$ is surjective this implies that ρ is surjective. The converse is trivial. □

10.1 Digression on Moduli

$X_0(N)(\overline{K})$ is the set of $\overline{K}$-isomorpisms classes of pairs (E, C) where $E/\overline{K}$ is an elliptic curve and C is a cyclic subgroup of order N. $X_0(N)(K)$ is the set of isomorphism classes of pairs (E, C) such that for all $\sigma \in \mathcal{G}al(\overline{K}/K)$, $\sigma(E, C) = (E, C)$. There is a map

$$\{k\text{-isomorphism classes of pairs } (E, C)/K\} \to X_0(N)(K)$$

which is "notoriously" non-injective. Deligne and Rapaport [13] prove the map is surjective. When $N = 1$ they observe that the map is surjective, then for $N > 1$ they show that certain obstructions vanish. A related question is

Question 10.1.1. *If $E/\overline{K}$ is isomorphic to all its Galois conjugates, is there E'/K which is isomorphic to E over $\overline{K}$?*

10.2 When is ρ_E Surjective?

Proposition 10.2.1. *Let E_1 and E_2 be elliptic curves defined over K with equal j-invariants, thus $E_1 \cong E_2$ over $\overline{K}$. Assume E_1 and E_2 do not have complex multiplication over $\overline{K}$. Then ρ_{E_1} is surjective iff ρ_{E_2} is surjective.*

Proof. Assume ρ_{E_1} is surjective. Since E_1 has no complex multiplication over $\overline{K}$, $\mathrm{Aut}\, E_1 = \{\pm 1\}$. Choose an isomorphism $\varphi : E_1 \xrightarrow{\sim} E_2$ over $\overline{K}$. Then for $\sigma \in \mathcal{G}al(\overline{K}/K)$ we have the diagram

$$\begin{array}{ccc} E_1 & \xrightarrow{\varphi} & E_2 \\ {\scriptstyle =}\downarrow & & {\scriptstyle =}\downarrow \\ {}^{\sigma}E_1 & \xrightarrow{\sigma\varphi} & {}^{\sigma}E_2 \end{array}$$

Thus $\sigma\varphi = \pm\varphi$ for all $\sigma \in \mathcal{G}al(\overline{K}/K)$, so $\varphi : E_1[N] \to E_2[N]$ defines an equivalence $\overline{\rho}_{E_1} \cong \overline{\rho}_{E_2}$. Since ρ_{E_1} is surjective this implies $\overline{\rho}_{E_2}$ is surjective which, by the previous proposition, implies ρ_{E_2} is surjective. ☐

Let $K = \mathbf{C}(j)$, with j transcendental over $\mathbf{C}$. Let E/K be an elliptic curve with j-invariant j. Fix a positive integer N and let

$$\rho_E : \mathcal{G}al(\overline{K}/K) \to \mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$$

be the associated Galois representation. Then $\det \rho_E$ is the cyclotomic character which is trivial since $\mathbf{C}$ contains the Nth roots of unity. Thus the image of ρ_E lands inside of $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$. Our next theorem states that a generic elliptic curve has maximal possible Galois action on its division points.

Theorem 10.2.2. $\rho_E : \mathcal{G}al(\overline{K}/K) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$ *is surjective.*

Igusa [8] found an algebraic proof of this theorem We will now make some comments on how an analytic proof goes.

Proof. Let $\mathbf{C}(j) = K = \mathcal{F}_1$ be the field of modular functions for $\mathrm{SL}_2(\mathbf{Z})$. Suppose $N \geq 3$ and let $\mathcal{F}_N$ be the field of mereomorphic functions for $\Gamma(N)$. Then $\mathcal{F}_N/\mathcal{F}_1$ is a Galois extension with Galois group $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$.

Let E be an elliptic curve over K with j-invariant j. We will show that $\mathcal{G}al(\mathcal{F}_N/\mathcal{F}_1) = \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$ acts transitively on the x-coordinates of the N torsion points of E. This will show that $\overline{\rho}_E$ maps surjectively onto $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$. Then by proposition 10.0.2, ρ_E maps surjectively onto $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$, as claimed.

We will now construct the x-coordinates of $E[N]$ as functions on $\mathcal{H}$ which are invariant under $\Gamma(N)$. (Thus $K(E[N]/\{\pm 1\}) \subset \mathcal{F}_N$.)

Let $\tau \in \mathcal{H}$ and let $L_\tau = \mathbf{Z}\tau + \mathbf{Z}$. Consider $\wp(z, L_\tau)$ which gives the x coordinate of $\mathbf{C}/L_\tau$ in it standard form. Define, for each nonzero $(r, s) \in ((\mathbf{Z}/N\mathbf{Z})/\{\pm 1\})^2$, a function

$$f_{(r,s)} : \mathcal{H} \to \mathbf{C} : \quad \tau \mapsto \frac{g_2(\tau)}{g_3(\tau)}\wp(\frac{r\tau + s}{N}, L_\tau).$$

First notice that for any $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$,

$$f_{(r,s)}(\alpha\tau) = f_{(r,s)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}(\tau).$$

Indeed, $\wp$ is homogenous of degree -2, g_2 is modular of weight 4 and g_3 is modular of weight 6, so

$$\begin{aligned} f_{(r,s)}(\alpha\tau) &= \frac{g_2(\alpha\tau)}{g_3(\alpha\tau)}\wp(\frac{r\alpha\tau + s}{N}) \\ &= (c\tau+d)^{-2}\frac{g_2(\tau)}{g_3(\tau)}\wp(\frac{ra\tau + rb + cs\tau + sd}{N(c\tau+d)}) \\ &= \frac{g_2(\tau)}{g_3(\tau)}\wp(\frac{(ra+sc)\tau + rb + sd}{N}) = f_{(r,s)\alpha}(\tau) \end{aligned}$$

If $\tau \in \mathcal{H}$ with $g_2(\tau), g_3(\tau) \neq 0$ then the $f_{(r,s)}(\tau)$ are the x-coordinates of the nonzero N-division points of $E_{j(\tau)}$. The various $f_{(r,s)}(\tau)$ are distinct. Thus $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$ acts transitively on the $f_{(r,s)}$. The consequence is that the $N^2 - 1$ nonzero division points of E_j have x-coordinates in $\overline{\mathcal{F}}_N$ equal to the $f_{(r,s)} \in \mathcal{F}_N$. □

10.3 Observations

Proposition 10.3.1. *If $E/\mathbf{Q}(\boldsymbol{\mu}_N)(t)$ is an elliptic curve with $j(E) = t$, then ρ_E has image* $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$.

Proof. Since $\mathbf{Q}(\boldsymbol{\mu}_N)$ contains the Nth roots of unity Nth cyclotomic character is trivial hence the determinent of ρ_E is trivial. Thus the image of ρ_E lies in $\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$. In the other direction, there is a natural inclusion

$$\mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z}) = \mathcal{G}al(\mathbf{C}(t)(E[N])/\mathbf{C}(t)) \hookrightarrow \mathcal{G}al(\mathbf{Q}(\boldsymbol{\mu}_N)(t)(E[N])/\mathbf{Q}(\boldsymbol{\mu}_N)(t)).$$

□

Proposition 10.3.2. *If $E/\mathbf{Q}(t)$ is an elliptic curve with $j(E) = t$, then ρ_E has image* $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$ *and* $\overline{\mathbf{Q}} \cap \mathbf{Q}(t)(E[N]) = \mathbf{Q}(\boldsymbol{\mu}_N)$.

Proof. Since $\mathbf{Q}(t)$ contains no Nth roots of unity, the mod N cyclotomic character, and hence $det\rho_E$, is surjective onto $(\mathbf{Z}/N\mathbf{Z})^*$. Since the image of ρ_E already contains $SL_2(\mathbf{Z}/N\mathbf{Z})$ it must equal $GL_2(\mathbf{Z}/N\mathbf{Z})$. For the second assertion consider the diagram

$$\begin{array}{ccccc} & \overline{\mathbf{Q}} & & \mathbf{Q}(t)(E[N]) & \\ & | & & | & SL_2 \\ & \mathbf{Q}(\boldsymbol{\mu}_N) & \subset & \mathbf{Q}(\boldsymbol{\mu}_N)(t) & \\ (\mathbf{Z}/N\mathbf{Z})^* & | & & | & GL_2/SL_2 = (\mathbf{Z}/N\mathbf{Z})^* \\ & \mathbf{Q} & \subset & \mathbf{Q}(t) & \end{array}$$

□

This gives a way to view $X_0(N)$ as a projective algebraic curve over $\mathbf{Q}$. Let $K = \mathbf{Q}(t)$ and let $L = K(E[N]) = \mathbf{Q}(\boldsymbol{\mu}_N)(t)$. Then

$$H = \{\left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right)\} \subset \mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z}) = \mathcal{G}al(L/K).$$

The fixed field L^H is an extension of $\mathbf{Q}(t)$ of transcendence degree 1 with field of constants $\overline{\mathbf{Q}} \cap L^H = \mathbf{Q}$, i.e., a projective algebraic curve.

10.4 A Descent Problem

Consider the following exercise which may be approached in an honest or dishonest way.

Exercise 10.4.1. Suppose L/K is a finite Galois extension and $G = \mathcal{G}al(L/K)$. Let E/L be an elliptic curve, assume $\operatorname{Aut}_L E = \{\pm 1\}$, and suppose that for all $g \in G$, there is an isomorphism ${}^gE \xrightarrow{\sim} E$ over L. Show that there exists E_0/K such that $E_0 \cong E$ over L.

Caution! The exercise is *false* as stated. Both the dishonet and honest approaches below work only if L is a separable closure of K. Now: can one construct a counterexample?

Discussion. First the hard, but "honest" way to look at this problem. For notions on descent see Serre [23]. By descent theory, to give E_0 is the same as to give a family $(\lambda_g)_{g\in G}$ of maps $\lambda_g : {}^gE \xrightarrow{\sim} E$ such that $\lambda_{gh} = \lambda_g \circ {}^g\lambda_h$ where ${}^g\lambda_h = g \circ \lambda_h \circ g^{-1}$. Note that $\lambda_g \circ {}^g\lambda_h$ maps ${}^{gh}E \to E$. This is the natural condition to impose, because if $f : E_0 \xrightarrow{\sim} E$ and we let $\lambda_g = f \circ {}^g(f^{-1})$ then $\lambda_{gh} = \lambda_g \circ {}^g\lambda_h$.

Using our hypothesis choose, for each $g \in G$, an isomorphism

$$\lambda_g : {}^gE \xrightarrow{\sim} E.$$

Define a map c by

$$c(g, h) = \lambda_g \circ {}^g\lambda_h \circ \lambda_{gh}^{-1}.$$

Note that $c(g, h) \in \operatorname{Aut} E = \{\pm 1\}$ so c defines an element of

$$H^2(G, \{\pm 1\}) \subset H^2(\mathcal{G}al(\overline{L}/K), \{\pm 1\}) = Br(K)[2].$$

Here $Br(K)[2]$ denotes the 2-torsion of the Brauer group

$$Br(K) = H^2(\mathcal{G}al(\overline{L}/K), \overline{L}^*).$$

This probably leads to an honest proof.

The dishonest approach is to note that $g(j(E)) = j(E)$ for all $g \in G$ since all conjugates of E are isomorphic and $j({}^gE) = g(j(E))$. Thus $j(E) \in K$ so we can define E_0/K by substituting $j(E)$ into the universal elliptic curve formula (see III.1.4 of [33]). This gives an elliptic curve E_0 defined over K but isomorphic to E over $\overline{K}$.

10.5 Second Look at the Descent Exercise

Last time we talked about the following problem. Suppose L/K is a Galois extension with $\operatorname{char} K = 0$, and let E/L be an elliptic curve. Suppose that for all $\sigma \in G = \mathcal{G}al(L/K)$, ${}^\sigma E \cong E$ over L. Conclude that there is an elliptic curve E_0/K such that $E_0 \cong E$ over L. The conclusion may fail to hold if L is a finite extension of K, but the exercise is true when $L = \overline{K}$. First we give a descent argument which holds when $L = \overline{K}$ and then give a counterexample to the more general statement.

For $g, h \in G = \mathcal{G}al(L/K)$ we define an automorphism $c(g, h) \in \operatorname{Aut} E = \{\pm 1\}$. Choose for every $g \in \mathcal{G}al(L/K)$ some isomorphism

$$\lambda_g : {}^gE \xrightarrow{\sim} E.$$

If the λ_g were to all satisfy the compatibility criterion $\lambda_{gh} = \lambda_g \circ {}^g\lambda_h$ then by descent theory we could find a K-structure on E, that is a model for E defined over K and isomorphic to E over L. Define $c(g, h)$ by $c(g, h)\lambda_{gh} = \lambda_g \circ {}^g\lambda_h$ so $c(g, h)$ measures how much the λ_g

fail to satisfy the compatibility criterion. Since $c(g, h)$ is a cocycle it defines an elements of $H^2(G, \{\pm 1\})$. We want to know that this element is trivial. When $L = \overline{K}$, the map $H^2(G, \{\pm 1\}) \to H^2(G, L^*)$ is injective. To see this first consider the exact sequence

$$0 \to \{\pm 1\} \to \overline{K}^* \xrightarrow{2} \overline{K}^* \to 0$$

where $2 : \overline{K}^* \to \overline{K}^*$ is the squaring map. Taking cohomology yields an exact sequence

$$H^1(G, \overline{K}^*) \to H^2(G, \{\pm 1\}) \to H^2(G, \overline{K}^*).$$

By Hilbert's theorem 90 ([25] Ch. X, Prop. 2), $H^1(G, \{\pm 1\}) = 0$. Thus we have an exact sequence

$$0 \to H^2(G, \{\pm 1\}) \to H^2(G, \overline{K}^*)[2] \to 0.$$

Thus $H^2(G, \{\pm 1\})$ naturally sits inside $H^2(G, L^*)$.

To finish Ribet does something with differentials and $H^0({}^gE, \Omega^1)$ which I don't understand.

The counterexample in the case when L/K is finite was provided by Kevin Buzzard (who said Coates gave it to him). Let $L = \mathbf{Q}(i)$, $K = \mathbf{Q}$ and E be the elliptic curve with Weirstrass equation $iy^2 = x^3 + x + 1$. Then E is isomorphic to its conjugate over L but one can show directly that E has no model over $\mathbf{Q}$.

10.6 Action of GL_2

Let $N > 3$ be an integer and $E/\mathbf{Q}(j)$ an elliptic curve with j-invariant $j(E) = j$. Then there is a Galois extension

$$\begin{array}{c} \mathcal{F}_N = \mathbf{Q}(j)(E[N]/\{\pm 1\}) \\ | \\ \mathcal{F}_1 = \mathbf{Q}(j) \end{array}$$

with Galois group $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})/\{\pm 1\}$. Think of $\mathbf{Q}(j)(E[N]/\{\pm 1\}$ as the field obtained from $\mathbf{Q}(j)$ by adjoining the x-coordinates of the N-torsion points of E. Note that this situation differs from the previous situation in that the base field $\mathbf{C}$ has been replaced by $\mathbf{Q}$.

Consider

$$\mathcal{F} = \bigcup_N \mathcal{F}_N$$

which corresponds to a projective system of modular curves . Let $\mathbf{A}_f$ be the ring of finite adèles, thus

$$\mathbf{A}_f = \hat{\mathbf{Q}} = \hat{\mathbf{Z}} \otimes \mathbf{Q} \subset \prod_p \mathbf{Q}_p.$$

$\mathbf{A}_f$ can be thought of as

$$\{(x_p) : x_p \in \mathbf{Z}_p \text{ for almost all } p\}.$$

The group $\mathrm{GL}_2(\mathbf{A}_f)$ acts on $\mathcal{F}$. To understand what this action is we first consider the subgroup $\mathrm{GL}_2(\hat{\mathbf{Z}})$ of $\mathrm{GL}_2(\mathbf{A}_f)$.

It can be shown that

$$\mathcal{F} = \mathbf{Q}(f_{N,(r,s)} : (r, s) \in (\mathbf{Z}/N\mathbf{Z})^2 - \{(0, 0)\}, N \geq 1)$$

where $f_{N,(r,s)}$ is a function

$$f_{N,(r,s)} : \mathcal{H} \to \mathbf{C} : \quad \tau \mapsto \frac{g_2(\tau)}{g_3(\tau)} \wp(\frac{r\tau + s}{N}, L_\tau).$$

We define the action of $\mathrm{GL}_2(\hat{\mathbf{Z}})$ on $\mathcal{F}$ as follows. Let $g \in \mathrm{GL}_2(\hat{\mathbf{Z}})$, then to give the action of g on $f_{N,(r,s)}$ first map g into $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$ via the natural reduction map, then note that $\mathrm{GL}_2(\mathbf{Z}/N\mathbf{Z})$ acts on $f_{N,(r,s)}$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot f_{N,(r,s)} = f_{N,(r,s)\begin{pmatrix} a & b \\ c & d \end{pmatrix}} = f_{N,(ra+sc,rb+sd)}.$$

Let E be an elliptic curve. Then the universal Tate module is

$$T(E) = \varinjlim E[N] = \prod_p T_p(E).$$

There is an isomorphism $\alpha : \hat{\mathbf{Z}}^2 \xrightarrow{\sim} T(E)$. Via right composition $\mathrm{GL}_2(\hat{\mathbf{Z}})$ acts on the collection of all such isomorphism α. So $\mathrm{GL}_2(\hat{\mathbf{Z}})$ acts naturally on pairs (E, α) but the action does nothing to E. One of the first important things we must do in understanding the construction of things like Shimura varieties is to free ourselves and allow $\mathrm{GL}_2(\hat{\mathbf{Z}})$ to act on the E's as well.

Let

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathbf{Q})$$

(thus g has positive determinent). Let $\tau \in \mathcal{H}$ and let $E = E_\tau$ be the elliptic curve determined by the lattice $L_\tau = \mathbf{Z} + \mathbf{Z}\tau$. Let

$$\alpha_\tau : L_\tau = \mathbf{Z} + \mathbf{Z}\tau \xrightarrow{\sim} \mathbf{Z}^2$$

be the isomorphism defined by $\tau \mapsto (1, 0)$ and $1 \mapsto (0, 1)$. Now view $\alpha = \alpha_\tau$ as a map

$$\alpha : \mathbf{Z}^2 \xrightarrow{\sim} H_1(E(\mathbf{C}), \mathbf{Z}).$$

Tensoring with $\mathbf{Q}$ then gives another map (also denoted α)

$$\alpha : \mathbf{Q}^2 \xrightarrow{\sim} H_1(E, \mathbf{Q}).$$

Then $\alpha \circ g$ is another isomorphism

$$\mathbf{Q}^2 \xrightarrow{\alpha \circ g} H_1(E, \mathbf{Q})$$

which induces an isomorphim $\mathbf{Z}^2 \xrightarrow{\sim} L' \subset H_1(E, \mathbf{Q})$ where L' is a lattice. There exists an elliptic curve $E'/\mathbf{C}$ and a map $\lambda \in \mathrm{Hom}(E', E) \otimes \mathbf{Q}$ which induces a map (also denoted λ)

$$\lambda : H_1(E', \mathbf{Z}) \xrightarrow{\sim} L' \subset H_1(E, \mathbf{Q})$$

on homology groups.

Now we can define an action on pairs (E, α) by sending (E, α) to (E', α'). Here α' is the map $\alpha' : \mathbf{Z}^2 \to H_1(E^1, \mathbf{Z})$ given by the composition

$$\mathbf{Z}^2 \xrightarrow{\alpha g} L' \xrightarrow{\lambda^{-1}} H_1(E^1, \mathbf{Z}).$$

In more concrete terms the action is

$$g : (E_\tau, \alpha_\tau) \mapsto (E'_\tau, \alpha'_\tau)$$

where $\tau' = g\tau = \frac{a\tau+b}{c\tau+d}$. [[CHECK THIS SOMETIME SOON!!]]

Chapter 11

Hecke Operators as Correspondences

11.1 Some Philosophy

We are studying modular forms over $\mathbf{C}$ and more generally over subrings R of $\mathbf{C}$. The Hecke algebras occur naturally as operators on various spaces of modular forms. We are aiming for an arithmetic perspective. One way is to study the arithmetic of cusp forms of weight 2 for congruence subgroups like $\Gamma_0(N)$ or $\Gamma_1(N)$. These cusp forms correspond to differentials on the modular curves $X_0(N)$ and $X_1(N)$. We have constructed models for each of these over $\mathbf{Q}$.

When $N'|N$ there is a natural map $X(N) \to X(N')$. Thus we get a tower of curves and a corresponding tower of number fields.

$$\begin{array}{cc}
\cdots & \cdots \\
\downarrow & \cup \\
X(N) & \mathcal{F}_N \\
\downarrow & \cup \\
X(N') & \mathcal{F}'_N \\
\downarrow & \cup \\
\cdots & \cdots
\end{array}$$

Taking limits gives a curve $X = \varprojlim X(N)$ and a corresponding field $\mathcal{F} = \varinjlim \mathcal{F}_N$.

There is an action of $\mathrm{GL}_2(\mathbf{A}_f)$ on pairs (E, α). By $\mathbf{A}_f$ we mean the ring of finite adèles, which may be identified with the restricted product $\prod_p \mathbf{Q}_p$. The subscript f, for 'finite', indicates that the infinite place is omitted. The full ring of adèles is $\mathbf{A} = \mathbf{A}_f \times \mathbf{R}$. If $g \in \mathrm{GL}_2(\mathbf{A}_f)$, then g acts on pairs (E, α) where E is an elliptic curve and

$$\alpha : \hat{\mathbf{Z}}^2 \xrightarrow{\sim} T(E) = \prod_p T_p(E).$$

Note that

$$\mathbf{A}_f = \hat{\mathbf{Z}} \otimes_{\mathbf{Z}} \mathbf{Q} = \prod \mathbf{Z}_p \otimes_{\mathbf{Z}} \mathbf{Q} = \hat{\mathbf{Q}},$$

and $T(E)$ is free of rank 2 over $\hat{\mathbf{Z}}$. Let

$$V(E) = T(E) \otimes_{\mathbf{Z}} \mathbf{Q} = \prod_p V_p(E)$$

where the product is restricted and $V_p(E) = T_p(E) \otimes_{\mathbf{Z}_p} \mathbf{Q}_p$. View $g \in \mathrm{GL}_2(\mathbf{A}_f)$ as an automorphism of $\hat{\mathbf{Q}}^2$. Then $\alpha \circ g$ sends $\hat{\mathbf{Z}}^2 \subset \hat{\mathbf{Q}}^2$ to a lattice $T' \subset V(E)$. As a lemma, one shows that there is an elliptic curve E' and a canonical map $\lambda : E' \to E$ such that the induced map $\lambda' : V(E') \xrightarrow{\sim} V(E)$ is an isomorphism which sends $T(E')$ maps to T' in $V(E)$. Then g sends the pair (E, α) to $(E', \lambda'^{-1} \circ \alpha \circ g)$.

11.2 Hecke Operators as Correspondences

Our goal is to think of Hecke operators (T_n, $\langle d \rangle$) as objects defined over $\mathbf{Q}$. We will define the Hecke operators as correspondences.

Definition 11.2.1. Let C_1 and C_2 be curves, then a **correspondence** $C_1 \to\to C_2$ is a curve C together with morphisms $\alpha : C \to C_1$ and $\beta : C \to C_2$.

$$\begin{tikzcd} & C \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ C_1 & & C_2 \end{tikzcd}$$

Giving a correspondence $C_1 \to\to C_2$ the **dual correspondence** $C_2 \to\to C_1$ is obtained by looking at the diagram in a mirror

$$\begin{tikzcd} & C \arrow[dl, "\beta"'] \arrow[dr, "\alpha"] & \\ C_2 & & C_1 \end{tikzcd}$$

The simplest case to consider is the modular curve $X_0(N)$ and Hecke operator T_p, where $p \nmid N$. We view T_p as a correspondence $X_0(N) \to\to X_0(N)$, thus there is a curve $C = X_0(pN)$ plus 2 maps α and β

$$\begin{tikzcd} & X_0(pN) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X_0(N) & & X_0(N) \end{tikzcd}$$

The maps α and β are degeneracy maps which forget data. To define them view $X_0(N)$ as classifying pairs (E, C) where E is an elliptic curve and $C \cong \mathbf{Z}/N\mathbf{Z}$ is a cyclic subgroup of order N. Similarly $X_0(pN)$ classifies pairs (E, G) where $G = C \oplus D \cong \mathbf{Z}/N\mathbf{Z} \oplus \mathbf{Z}/p\mathbf{Z}$ and D is cyclic of order p. Then

$$\begin{aligned} \alpha &: (E, G) \mapsto (E, C) \\ \beta &: (E, G) \mapsto (E/D, (C + D)/D) \end{aligned}$$

Now we now translate this into the language of complex analysis. The first map α corresponds to the map

$$\Gamma_0(pN)\backslash\mathcal{H} \to \Gamma_0(N)\backslash\mathcal{H}$$

induced by the inclusion $\Gamma_0(pN) \hookrightarrow \Gamma_0(N)$. The second map β is constructed by composing the map

$$\Gamma_0\backslash\mathcal{H} \xrightarrow{\sim} \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \Gamma_0(pN) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}^{-1} \backslash\mathcal{H}$$

with the map to $\Gamma_0(N)\backslash\mathcal{H}$ induced by the inclusion

$$\Gamma_0(N) \subset \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \Gamma_0(pN) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}^{-1}$$

The maps α and β induce maps

$$\alpha^*, \beta^* : H^0(X_0(N), \Omega^1) \to H^0(X_0(pN), \Omega^1).$$

We can identify $S_2(\Gamma_0(N))$ with $H^0(X_0(N), \Omega^1$ by sending the cusp form $f(\tau)$ to the holomorphic differential $f(\tau)d\tau$. Now we get maps

$$\alpha^*, \beta^* : S_2(\Gamma_0(N)) \to S_2(\Gamma_0(pN)).$$

Exercise 11.2.2. Show that $\alpha^*(f) = f$ thought of as a cusp form with respect to the smaller group $\Gamma_0(pN)$. Then show that

$$\beta^*(f) = p \sum_{n=1}^{\infty} a_n q^{pN}.$$

(Ribet was unsure whether the factor of p should be there.)

11.3 Generalities on Correspondences

Let X, Y, and C be curves and let α and β be nonconstant holomorphic maps so that we have the correspondence

$$\begin{tikzcd} & C \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X & & Y \end{tikzcd}$$

Then we have maps

$$H^0(X, \Omega^1) \xrightarrow{\alpha^*} H^0(C, \Omega^1) \xrightarrow{\beta_*} H^0(Y, \Omega^1).$$

The composition $\beta_* \circ \alpha^*$ is a map

$$H^0(X, \Omega^1) \to H^0(Y, \Omega^1).$$

Switching the roles of X and Y gives a map

$$H^0(Y, \Omega^1) \to H^0(X, \Omega^1).$$

In this context we can identify T_p by viewing it as the map

$$\beta_* \circ \alpha^* : H^0(X_0(N), \Omega^1) \to H^0(X_0(N), |Omega^1)$$

and using the fact that $S_2(\Gamma_0(N)) = H^0(X_0(N), \Omega^1)$. One should recover the explicit representation

$$T_p : \sum a_n q^n \mapsto \sum a_{np} q^n + p \sum a_n q^{np}.$$

Now lets think more generally about correspondences. Suppose $\varphi : X \to Y$ is a map of curves. Let $\Gamma \subset X \times Y$ be the graph of φ. This gives a stupid correspondence

$$\begin{tikzcd} & \Gamma \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X & & Y \end{tikzcd}$$

We can reconstruct φ since $\varphi(x) = \beta(\alpha^{-1}(x))$.

More generally suppose $\alpha : \Gamma \to X$ has degree $d \geq 1$. View $\alpha^{-1}(x)$ as a divisor on Γ. Then $\beta(\alpha^{-1})$ is a divisor on Y. We thus get a map

$$\operatorname{Div}^n X \xrightarrow{\beta\circ\alpha^{-1}} \operatorname{Div}^{dn} Y.$$

In particular, when $d = 0$, there we get a map $\operatorname{Div}^0 X \to \operatorname{Div}^0 Y$. Now apply this to the special case of T_p. T_p is the correspondence

$$\begin{tikzcd} & X_0(pN) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X_0(N) & & X_0(N) \end{tikzcd}$$

and the induced map is

$$(E, C) \overset{\alpha^*}{\mapsto} \sum_{D\in E[p]} (E, C \oplus D) \overset{\beta}{\mapsto} \sum_{D\in E[p]} (E/D, (C + D)/D)$$

Thus we have a map

$$\left\{ \text{ divisors on } X_0(N) \right\} \longrightarrow \left\{\text{divisors on } X_0(N)\right\}$$

This strongly resembles the original definition of T_p as a correspondence of lattices. But there was a factor of p^{k-1} then which annoyed Ogus. It disappears here though, and this is related to the 'p' in the exercise.

11.4 Jacobians of Curves

Let X be a curve of genus g over a field k. There is an important association

$$\left\{ \text{ curves } X/k \right\} \longrightarrow \left\{ \text{ Jacobians } Jac(X) = J(X) \text{ of curves } \right\}$$

between curves and their associated Jacobians.

Definition 11.4.1. Let X be a curve, then the **Jacobian** of X is an abelian variety of dimension g whose underlying group is isomorphic to the group of divisors of degree 0 on X modulo linear equivalence.

There are many constructions of the Jacobian of a curve. We first consider the Albanese construction. Recall that over $\mathbf{C}$ an abelian variety is just $\mathbf{C}^g/\mathcal{L}$ where $\mathcal{L}$ is a lattice (and hence a free $\mathbf{Z}$-module of rank $2g$). There is an embedding

$$H_1(X, \mathbf{Z}) \hookrightarrow H^0(X, \Omega^1)^*$$
$$\gamma \mapsto \int_\gamma \bullet$$

Then we realize $J(X)$ as a quotient

$$J(X) = H^0(X, \Omega^1)^*/H_1(X, \mathbf{Z}).$$

In this construction $J(X)$ is covariantly associated to X. There are other constructions in which $J(X)$ is contravariantly associated to X. If F is a correspondence $X \to\to Y$, then F induces a map $J(X) \to J(Y)$ and also a map $J(Y) \to J(X)$. If $X = Y$ so that X and Y are the same, one can often be confused which duality should be used. Fortunately, for T_p when p is prime to N it does not matter. But it matters a lot if $p|N$ since then we have noncommuting confusable operators and, "this has resulted in lots of mistakes in the literature."

11.5 More on Hecke Operators

Our goal is to move things down to $\mathbf{Q}$ from $\mathbf{C}$ or $\overline{\mathbf{Q}}$. In doing this we want to understand T_n (or T_p), that is, how they act on the associated Jacobians and how they can be viewed as correspondences. In characteristic p the formulas of Eichler-Shimura will play an important role.

We consider T_p as a correspondence on $X_1(N)$ or $X_0(N)$. To avoid confusion we will mainly consider T_p on $X_0(N)$ with $p\nmid N$. Thus assume, unless otherwise stated, that $p\nmid N$. Remember that T_p was defined to be the correspondence

$$\begin{array}{ccc} & X_0(pN) & \\ {}^{\alpha}\swarrow & & \searrow^{\beta} \\ X_0(N) & & X_0(N) \end{array}$$

Think of $X_0(pN)$ as consisting of pairs $(\underline{E}, D)$ where D is a cyclic subgroup of E of order p and $\underline{E}$ is the *enhanced* elliptic curve consisting of an elliptic curve E along with a cyclic subgroup of order N. The degeneracy map α forgets the subgroup D and the degeneracy map β divides by it. By contravariant functoriality we have a commutative diagram

$$\begin{array}{ccc} H^0(X_0(N),\Omega^1) & \xrightarrow{T_p^*=\alpha_*\circ\beta^*} & H^0(X_0(N),\Omega^1) \\ \Vert & & \Vert \\ S_2(\Gamma_0(N)) & \xrightarrow{T_p} & S_2(\Gamma_0(N)) \end{array}$$

Our convention to define T_p^* as $\alpha_*\circ\beta^*$ instead of $\beta_*\circ\alpha^*$ was completely psychological because there is a canonical duality relating the two. We chose the way we did because of the analogy with the case of a morphism $\varphi : Y \to X$ with graph Γ which induces a correspondence

$$\begin{array}{ccc} & \Gamma & \\ {}^{\pi_1}\swarrow & & \searrow^{\pi_2} \\ Y & & X \end{array}$$

Since the morphism φ induces a map on global sections in the other direction

$$H^0(X,\Omega^1)\Gamma(X) \xrightarrow{\varphi^*} \Gamma(Y) = H^0(Y,\Omega^1)$$

it is psychologically natural for more general correspondence such as T_p to map from the right to the left.

The morphisms α and β in the definition of T_p are defined over $\mathbf{Q}$. This can be seen using the general theory of representable functors [How?]. Thus since T_p is defined over $\mathbf{Q}$ most of the algebraic geometric objects we will construct related to T_p will be defined over $\mathbf{Q}$.

11.6 Hecke Operators acting on Jacobians

The Jacobian $J(X_0(N)) = J_0(N)$ is an abelian variety defined over $\mathbf{Q}$. There are both covariant and contravariant formations of $J_0(N)$. Thus a map $\alpha : X_0(pN) \to X_0(N)$ induces maps

$$\begin{array}{ccc} J_0(pN) & & J_0(pN) \\ \alpha^*\uparrow & & \downarrow\alpha_* \\ J_0(N) & & J_0(N) \end{array}$$

Note that $\alpha_* \circ \alpha^* : J_0(N) \to J_0(N)$ is just multiplication by $\deg(\alpha) = p+1$, since there are $p+1$ subgroups of order p in $\underline{E}$. (At least when $p \nmid N$, when $p|N$ there are only p subgroups.)

We will often refer to T_p as ξ_p as Shimura does in his book. There are two possible ways to define $T_p = \xi_p$ as an endomorphism of $J_0(N)$. We could either define T_p as $\beta_* \circ \alpha^*$ or equivalently as $\alpha_* \circ \beta^*$ (assuming still that $p \nmid N$).

11.6.1 The Albanese Map

There is a way to map the curve $X_0(N)$ into its Jacobian since the underlying group structure of $J_0(N)$ is

$$J_0(N) = \frac{\left\{ \text{divisors of degree 0 on } X_0(N) \right\}}{\left\{ \text{principal divisors} \right\}}$$

Once we have chosen a rational point, say ∞, on $X_0(N)$ we obtain the Albanese map

$$\theta : X_0(N) \to J_0(N) : x \mapsto x - \infty$$

which sends a point x to the divisor $x - \infty$. The map θ gives us a way to pullback differentials on $J_0(N)$. Let $\operatorname{Cot} J_0(N)$ denote the cotangent space of $J_0(N)$ (or the space of regular differentials). The diagram

$$\begin{array}{ccc} \operatorname{Cot} J_0(N) & \xleftarrow{\xi_p^*} & \operatorname{Cot} J_0(N) \\ \theta^* \big\downarrow \wr & & \wr \big\downarrow \theta^* \\ H^0(X_0(N), \Omega^1) & \xleftarrow{T_p^*} & H^0(X_0(N), \Omega^1) \end{array}$$

may be taken to give a definition of ξ_p since there is a unique endomorphism $\xi_p : J_0(N) \to J_0(N)$ inducing a map ξ_p^* which makes the diagram commute.

Now suppose Γ is a correspondence $X \to\to Y$ so we have a diagram

$$\begin{array}{ccc} & \Gamma & \\ {}^{\alpha}\swarrow & & \searrow^{\beta} \\ X & & Y \end{array}$$

For example, think of Γ as the graph of a morphism $\varphi : X \to Y$. Γ should naturally induce a map

$$H^0(Y, \Omega^1) \longrightarrow H^0(X, \Omega^1).$$

Taking Jacobians we see that the composition

$$J(X) \xrightarrow{\alpha^*} J(\Gamma) \xrightarrow{\beta_*} J(Y)$$

gives a map $\beta_* \circ \alpha^* : J(X) \to J(Y)$. On cotangent spaces this induces a map

$$\alpha^* \circ \beta_* : H^0(Y, \Omega^1) \to H^0(X, \Omega^1).$$

Now, after choice of a rational point, the map $X \to J(X)$ induces a map $\operatorname{Cot} J(X) \to H^0(X, \Omega^1)$. This is in fact independent of the choice of rational point since differentials on $J(X)$ are invariant under translation.

The map $J(X) \to J(Y)$ is preferred in the literature. It is said to be induced by the Albanese functoriality of the Jacobian. We could have just as easily defined a map from $J(Y) \to J(X)$. To see this let

$$\psi = \beta_* \circ \alpha^* : J(X) \to J(Y).$$

Dualizing induces a map $\psi^\vee = \alpha_* \circ \beta^*$

$$\begin{array}{ccc} J(X)^\vee & \xleftarrow{\psi^\vee} & J(Y)^\vee \\ \wr\Big| & & \wr\Big| \\ J(X) & & J(Y) \end{array}$$

where we have used the fact that Jacobians are canonically autodual. This canonical duality is discussed in Mumford's [17] and [18] and in Milne's article in [30].

11.6.2 The Hecke Algebra

We now have $\xi_p = T_p \in \operatorname{End} J_0(N)$ for every prime p. If $p|N$, then we must decide between $\alpha_* \circ \beta^*$ and $\beta_* \circ \alpha^*$. The usual choice is the one which induces the usual T_p on cusp forms. If you don't like your choice you can get out of it with Atkin-Lehrer operators.

Let

$$\mathbf{T} = \mathbf{Z}[\ldots, T_p, \ldots] \subset \operatorname{End} J_0(N)$$

then $\mathbf{T}$ is the same as $\mathbf{T}_\mathbf{Z} \subset \operatorname{End}(S_2(\Gamma_0(N)))$. To see this first note that there is a map $\mathbf{T} \to \mathbf{T}_\mathbf{Z}$ which is not a prior injective, but which is injective because elements of $\operatorname{End} J_0(N)$ are completely determined by their action on $\operatorname{Cot} J_0(N)$.

11.7 The Eichler-Shimura Relation: Part I

$J_0(N)$ is an abelian variety defined over $\mathbf{Q}$ so it is natural to ask for those primes p for which $J_0(N)$ have good reduction. In the 1950's Igusa managed to show that $J_0(N)$ has good reduction for all $p \nmid N$. He did this by first viewing $J_0(N)$ as a scheme over $\operatorname{spec} \mathbf{Q}$, then "spreading things out" to make an abelian scheme over $\operatorname{spec} \mathbf{Z}[1/N]$. He did this by taking the Jacobian of the normalization of $X_0(N)$ (which is defined over $\mathbf{Z}[1/N]$) in $\mathbf{P}_{\mathbf{Z}[1/N]}$.

The Eichler-Shimura formula is a formula for T_p in characteristic p (or equivalently for endomorphisms $\xi_p \in \operatorname{End}(J_0(N)/\mathbf{F}_p)$) for all p for which $J_0(N)$ has good reduction at p. Suppose now that $p \nmid N$, then $X_0(N)/\mathbf{F}_p$ has the same sort of properties as $X_0(N)/\mathbf{Q}$. In particular it classifies enhanced elliptic curves $\underline{E}$ where $\underline{E} = (E, C)$ is an elliptic curve over $\mathbf{F}_p$ along with a cyclic subgroup of order N.

Next we ask what happens to the map $\xi_p : X_0(N) \to X_0(N)/\mathbf{Q}$ in reduction. Thus consider the correspondence

$$\begin{array}{ccc} & X_0(Np) & \\ \alpha \swarrow & & \searrow \beta \\ X_0(N) & & X_0(N) \end{array}$$

$X_0(N)$ has good reduction at p, but $X_0(Np)$ may not. Fortunately Deligne and Rapaport showed that $X_0(Np)$ has relatively benign reduction at p. The first step is to form a model of $X_0(Np)/\mathbf{F}_p$. Over $\mathbf{F}_p$, $X_0(Np)$ can be thought of as two copies of $X_0(N)$ glued by pairing off the supersingular points (thus corresponding supersingular points are glued). The set of supersingular points

$$\Sigma \subset X_0(N)(\overline{\mathbf{F}}_p)$$

is the set of points represented by supersingular elliptic curves. There are only finitely many of these, in fact $g+1$ where g is the genus of $X_0(N)$.

Next we must understand the following constellation of diagrams

$$\begin{tikzcd}
X_0(N) \arrow[dd] \arrow[dr] & & X_0(N) \arrow[dd] \arrow[dl] \\
& X_0(Np) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
X_0(N) & & X_0(N)
\end{tikzcd}$$

11.8 The Eichler-Shimura Relation: Part II

Suppose p is a prime, $p \nmid N$. Consider the correspondence $T_p : X_0(N) \to\to X_0(N)$ which takes an enhanced elliptic curve $\underline{E}$ to the sum $\sum_{\#D=p} \underline{E}/D$ of quotients of $\underline{E}$ by subgroups of order p. This is the correspondence

$$\begin{tikzcd}
& X_0(pN) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
X_0(N) & & X_0(N)
\end{tikzcd}$$

From this one gets $\xi_p = T_p : J_0(N) \to J_0(N)$ by functoriality. (Henceforth we will denote Shimura's ξ_p by T_p.) There are many ways to think of $J_0(N)$. The cotangent space $\mathrm{Cot}\, J_0(N)$ of $J_0(N)$ is the space of holomorphic (or translation invariant) differentials on $J_0(N)$, which is isomorphic to $S_2(\Gamma_0(N))$. This gives a connection between our geometric definition of T_p and the definition, presented at the beginning of the course, of T_p as an operator on a space of cusp forms.

The Eichler-Shimura relation takes place in $\mathrm{End}(J_0(N)_{\mathbf{F}_p})$. Since $X_0(N)$ reduces "nicely" in characteristic p, we can apply the Jacobian construction to $X_0(N)_{\mathbf{F}_p}$. Furthermore the natural reduction map

$$\mathrm{End}(J_0(N)) \hookrightarrow \mathrm{End}(J_0(N))_{\mathbf{F}_p}$$

is injective. [[Why is this true? It seems at first glance to be the sort of thing that might be false, so why is it true?]] Let $F : X_0(N)_{\mathbf{F}_p} \to X_0(N)_{\mathbf{F}_p}$ be the Frobenius map in characteristic p. Thus, if $K = K(X_0(N))$ is the function field of the nonsingular curve $X_0(N)$, then F is induced by the p-th power map $K \to K : a \mapsto a^p$. By both functorialities F induces maps on the Jacobian of $X_0(N)_{\mathbf{F}_p}$.

$$\begin{tikzcd}
J_0(N)_{\mathbf{F}_p} \arrow[r, shift left, "\mathrm{Ver}"] & J_0(N)_{\mathbf{F}_p} \arrow[l, shift left, "\mathrm{Frob}"]
\end{tikzcd}$$

Note that $\mathrm{Ver} \circ \mathrm{Frob} = \mathrm{Frob} \circ \mathrm{Ver} = p$ since p is the degree of F. Since the Albanese functoriality is given by pullback of differentials,

$$\mathrm{Frob} = F_* \quad \text{and} \quad \mathrm{Ver} = \mathrm{Frob}^\vee = F^*.$$

Theorem 11.8.1 (Eichler-Shimura). *The formula of Eichler-Shimura is*

$$T_p = \mathrm{Frob} + \mathrm{Ver}\,.$$

Proof. View $X_0(pN)_{\mathbf{F}_p}$ as two copies of $X_0(N)_{\mathbf{F}_p}$ glued along corresponding supersingular points. Now consider the diagram of schemes over $\mathbf{F}_p$.

$$\begin{tikzcd}
X_0(N)_{\mathbf{F}_p} \arrow[dr, "r"] & & X_0(N)_{\mathbf{F}_p} \arrow[dl, "s"'] \\
& X_0(pN)_{\mathbf{F}_p} \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
X_0(N)_{\mathbf{F}_p} & & X_0(N)_{\mathbf{F}_p}
\end{tikzcd}$$

We must define r and s. How can we define a map $X_0(N)_{\mathbf{F}_p} \to X_0(pN)_{\mathbf{F}_p}$ in characteristic p? A point of $X_0(N)_{\mathbf{F}_p}$ is an enhanced elliptic curve $\underline{E} = (E, C)$ consisting of an elliptic curve (not necessarily defined over $\mathbf{F}_p$) along with a cyclic subgroup of order N. The map r sends $\underline{E}$ to $\underline{E}$ plus an isogeny of degree p,

$$E \xrightarrow{\text{frob}} E^{(p)}$$

where $E^{(p)}$ is the curve obtained from E by hitting all defining equations by Frobenious, that is, by p-th powering the coefficients of the defining equations for E. We must introduce $E^{(p)}$ since E might not be defined over $\mathbf{F}_p$ so Frobenious wouldn't be an endomorphism of E. Thus r is the map

$$r: \quad \underline{E} \mapsto (\underline{E}, E \xrightarrow{\text{frob}} E^{(p)})$$

and similarly we define s to be the map

$$s: \quad \underline{E} \mapsto (\underline{E}, E \xleftarrow{\text{ver}} E^{(p)}$$

where ver is the dual of frob. We may view α as the map sending an isogeny $E \to E'$ to the curve E, and similarly we may view β as the map sending an isogeny $E' \to E$ to the curve E'. Viewing elements of $X_0(N)$ as isogenies of degree N is equivalent to our previous view of $X_0(N)$ as consisting of pairs (E, C) where C is a cyclic subgroup of order N, but the isogeny viewpoint is better suited to the purpose at hand. Thus

$$\begin{aligned} \alpha: &\quad (E \to E') \mapsto E \\ \beta: &\quad (E' \to E) \mapsto E' \end{aligned}$$

It now follows immediately that $\alpha \circ r = \mathrm{id}$ and $\beta \circ s = \mathrm{id}$. Note also that $\alpha \circ s = \beta \circ r = F$ is the map $E \mapsto E^{(p)}$.

At least intuitively we may view $X_0(pN)_{\mathbf{F}_p}$ as the disjoint union of two copies of $X_0(N)_{\mathbf{F}_p}$. Then

$$\begin{tikzcd}
& X_0(pN)_{\mathbf{F}_p} \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
X_0(N)_{\mathbf{F}_p} & & X_0(N)_{\mathbf{F}_p}
\end{tikzcd}
=
\begin{tikzcd}
& X_0(N)_{\mathbf{F}_p} \arrow[dl, "\mathrm{id}=\alpha\circ r"'] \arrow[dr, "F=\beta\circ r"] & \\
X_0(N)_{\mathbf{F}_p} & & X_0(N)_{\mathbf{F}_p}
\end{tikzcd}
+
\begin{tikzcd}
& X_0(N)_{\mathbf{F}_p} \arrow[dl, "F=\beta\circ r"'] \arrow[dr, "\mathrm{id}=\alpha\circ r"] & \\
X_0(N)_{\mathbf{F}_p} & & X_0(N)_{\mathbf{F}_p}
\end{tikzcd}$$

[[For some reason I guess this is supposed to prove that $T_p = \mathrm{Frob} + \mathrm{Ver}$. Why?]] □

11.9 Applications

We have the relation $T_p = \mathrm{Frob} + \mathrm{Ver}$ in $\mathrm{End}(J_0(N)_{\mathbf{F}_p})$. How can we apply this formula? Consider the ℓ-adic Tate module

$$\mathrm{Tate}_\ell(J_0(N)) = \varprojlim J_0(N)[\ell^\nu] \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$$

which is a vector space of dimension $2g(X_0(N))$ over $\mathbf{Q}_\ell$. When $\ell \neq p$ the image of $\mathrm{Tate}_\ell(J_0(N))$ under reduction mod p is basically the same as the original module in characteristic 0. Denote also by $\mathrm{Tate}_\ell(J_0(N))$ the image of $\mathrm{Tate}_\ell(J_0(N))$ under reduction mod p. On $\mathrm{Tate}_\ell(J_0(N))$ we have linear operators Frob_p, Ver_p and T_p which satisfy $T_p = \mathrm{Frob}_p + \mathrm{Ver}_p$ and

$$\mathrm{Frob}_p \circ \mathrm{Ver}_p = \mathrm{Ver}_p \circ \mathrm{Frob}_p = p = \text{multiplication by } p.$$

Since p is invertible (because p is prime to ℓ),

$$T_p = \mathrm{Frob}_p + p(Frob_p)^{-1}$$

and thus

$$\mathrm{Frob}_p^2 - T_p \mathrm{Frob}_p + p = 0.$$

This is a beautiful quadratic relation and so we should be able to get something out of it. We will come back to this shortly, but first we consider the various objects acting on the ℓ-adic Tate module. $\mathrm{Tate}_\ell(J_0(N))$ is acted upon in a natural way by

1. $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$
2. $\mathrm{End}_{\mathbf{Q}}(J_0(N)) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$ (which acts by functoriality)

These actions commute with each other since endomorphisms defined over $\mathbf{Q}$ aren't effected by the action of the Galois group of $\mathbf{Q}$. Reducing mod p we also have the commuting actions

3. $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$
4. $\mathrm{End}_{\mathbf{F}_p}(J_0(N)) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$

Note that a decomposition group group $D_p \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts (after modding out by the corresponding inertia group) in the same way as $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ and the action is unramified, thus 4 is a special case of 1. [[This could be wrong, because I'm perpetually confused about this idea.]]

The Frobenius elements $\mathrm{Frob} \in \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$ and $\mathrm{Frob} \in \mathrm{End}_{\mathbf{F}_p}(J_0(N)) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$ induce the *same* operator on Tate modules. In other words, on the Tate module the morphism Frob_p on the Jacobian acts in the same way as $\varphi_p \in \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$. Note that while φ is in a quotient of a decomposition group one often takes a lift to get an element $\varphi_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

On $\mathrm{Tate}_\ell(J_0(N))$ we have a quadratic relationship

$$\varphi_p^2 - T_p\varphi_p + p = 0.$$

This plays a role when one tries to separate out pieces of the Jacobian. Let

$$R = \mathbf{Z}[\ldots, T_p, \ldots] \subset \mathrm{End}\, J_0(N)$$

where we only adjoin those T_p with $p \not| N$. Think of R as a reduced Hecke algebra, in particular, R is a subring of $\mathbf{T}$. Then

$$R \otimes \mathbf{Q} = \prod_{i=1}^{r} E_i$$

where the E_i are totally real number fields. The factors E_i are in bijection with the Galois conjugacy classes of weight 2 newforms f on $\Gamma_0(M)$ (for some $M|N$). Galois acts on newforms by hitting their coefficients. The bijection is the map

$$f \mapsto \mathbf{Q}(\text{coefficients of } f) = E_i$$

Observe that the map is the same if we replace f by one of its conjugates. [[In order to see this don't we need to know that $\mathbf{Q}$(coeffics of f) is Galois? Do we know this?]] This decomposition is a decomposition of a subring

$$R \otimes \mathbf{Q} \subset \operatorname{End}(J_0(N)) \otimes \mathbf{Q} \stackrel{\text{def}}{=} \operatorname{End}(J_0(N) \otimes \mathbf{Q}).$$

Thus it induces a direct product decomposition of $J_0(N)$, so $J_0(N)$ gets divided up into subvarieties which correspond to conjugacy classes of newforms. A newform of exact level N corresponds to an abelian variety of dimension N.

The relationship

$$\varphi_p^2 - T_p\varphi_p + p = 0$$

reminds one of the Caley-Hamilton theorem in which the coefficients would be the trace and determinent. Thus it seems reasonable to expect that

$$\operatorname{Tr}\varphi_p = T_p \quad \text{and} \quad \det\varphi_p = p.$$

This is is true but it does not follow formally from the given quadratic relation. It is proved in a fairly direct manner using the Weil pairing.

11.10 More on Eichler-Shimura

Remark 11.10.1 (Useful References). At the time of this writing (March, 1996) several useful papers on last year's Fermat conference could be obtained via `ftp` from

`gauss.math.brown.edu:/dist/FermatConference`

Viewing T_p as an element of $\operatorname{End}_{\mathbf{F}_p}(J_0(N)_{\mathbf{F}_p})$ we have the Eichler-Shimura relation

$$T_p = \operatorname{Frob} + \operatorname{Ver}.$$

Let $\mathbf{T}$ be the algebra of Hecke operators viewed as a subring of $\operatorname{End}_{\mathbf{Q}}(J_0(N))$. Then since $S_2(\Gamma_0(N))$ is the tangent space to $J_0(N)$

$$\begin{aligned}\mathbf{T} &\subset \operatorname{End}_{\mathbf{Q}}(J_0(N)) \subset \operatorname{End}_{\mathbf{C}}(J_0(N)) \\ &\subset \operatorname{End}_{\mathbf{C}}(\operatorname{Hom}(S_2(\Gamma_0(N)), \mathbf{C}).\end{aligned}$$

[I don't understand how this is the tangent space!] Thus the $\mathbf{T}$ constructed before is really the same ring of Hecke operators as we obtained via the modular forms construction.

Let $X = \operatorname{Tate}_\ell(J_0(N))$ and suppose p is a prime different from ℓ. If g is the genus of $X_0(N)$ then X is a $\mathbf{Q}_\ell$-vector space of dimension $2g$. Since the ℓ-torsion of $J_0(N)_{\mathbf{F}_p}$ is the reduction mod p of the ℓ-torsion of $J_0(N)$ the operator T_p acts on X. Similarly if $\varphi_p \in \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is a Frobenius element then there is an associated map Frob_p on X. The Eichler-Shimura relation lifts to characteristic 0 to give [[how? why? this isn't clear.]

$$T_p = \varphi_p + p\varphi_p^{-1}.$$

Chapter 12

Abelian Varieties from Modular Forms

In this section we will describe how to break $J_0(N)$ up into subvarieties A_f corresponding to cusp forms f. This was originally done by Shimura, see Theorem 7.14 of [31]. Let N be a positive integer and let $\mathbf{T}$ be the algebra of Hecke operators on $J_0(N)$. Let $\mathbf{T}_0 = \mathbf{Z}[\dots, T_n, \dots : (n,N)=1] \subset \mathbf{T}$ be the subalgebra of $\mathbf{T}$ obtained by adjoining to $\mathbf{Z}$ only those Hecke operators T_n with n relatively prime to N. If $f = \sum a_n q^n$, let $E_f = \mathbf{Q}(a_1, a_2, \dots)$. Then as in the previous section

$$\mathbf{T}_0 \otimes \mathbf{Q} = \prod_f E_f,$$

where f runs over a set of representatives of the space of newforms of weight 2 and level dividing N modulo conjugation by $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. We call the conjugates of a newform its companions. Each f in the product $\mathbf{T}_0 \otimes \mathbf{Q}$ contains a projection $\pi_f = (0, \dots, 0, 1, 0, \dots, 0) \in \prod E_f$ where the 1 is in the fth position. Since $\mathbf{T}_0 \subset \mathbf{T}$ and $\mathbf{T}$ is free over $\mathbf{Z}$, these projectors lie in $\mathbf{T} \otimes \mathbf{Q}$. Since $\mathbf{T} \otimes \mathbf{Q}$ is commutative and the π_f are mutually orthogonal idempotents whose sum is $(1, 1, \dots, 1)$ we see that $\mathbf{T} \otimes \mathbf{Q}$ breaks into a product of algebras

$$\mathbf{T} \otimes \mathbf{Q} = \prod \mathscr{A}_f.$$

(The map is $x \mapsto \sum x\pi_f$.)

Exercise 12.0.2. Show that $\dim_{E_f} \mathscr{A}_f$ is the number of divisors of $\frac{N}{N(f)}$ where $N(f)$ is the level of the newform f.

Let's examine a particular case of this exercise. Let q be a prime, suppose f is a newform of level $N(f)$, and $N = N(f)q$ with $q \nmid N(f)$. We show that

$$\mathscr{A}_f = E_f[U]/(U^2 - a_q U + q)$$

and hence $\dim_{E_f} \mathscr{A}_f = 2$ which, as expected, is the number of divisors of $N/N(f) = q$. The first step is to view $\mathscr{A}_f$ as the space of operators generated by the T_p acting on the cusp forms $f(d\tau)$ for $d | (N/N(f))$. In our special case T_p acts as the scalar a_p when $p \nmid N$. When $q | N$, T_q acts as the operator often called U_q. Recall that U_q corresponds to the matrix $\begin{pmatrix} a_q & 1 \\ -q & 0 \end{pmatrix}$ with respect to a suitable basis (see section 9). Thus U_q satisfies the relation $U_q^2 - a_q U + q$.

Exercise 12.0.3. Write $\mathscr{A}_f$ in general as

$$E_f[\dots, U_q, \dots]/I$$

where I is an ideal and the U_q correspond to primes dividing $N/N(f)$. [Hint: See Lemma 4.4 of [4] for the answer.]

One can define an abelian subvariety A_f of $J_0(N)$ as follows. As above let π_f be the fth projector in $\mathbf{T}_0 \otimes \mathbf{Q} = \prod_g E_g$. We would like to define A_f to be the image of $J_0(N)$ under π_f. But π_f isn't an endomorphism of $J_0(N)$. But $\pi_f \in \mathrm{End}(J_0(N)) \otimes \mathbf{Q}$ so there exists n such that $n\pi_f \in \mathrm{End}(J_0(N))$. It thus makes sense to let $A_f = n\pi_f(J_0(N))$. This is independent, up to isogeny, of the choice of n.

Due to the fact that A_f is only defined up to isogeny it is natural to consider the category of abelian varieties up to isogeny (whose objects are abelian varieties and morphisms are isogenies over $\mathbf{Q}$). In this category $J_0(N)$ is decomposes into subvarieties A_f since we have an isogeny

$$\prod_f A_f \to J_0(N).$$

Over $\mathbf{C}$ we can view this decomposition in yet another way. Think of $J_0(N)_{\mathbf{C}}$ as

$$J_0(N) = \mathrm{Hom}(S_2(\Gamma_0(N)), \mathbf{C})/H_1(X_0(N), \mathbf{Z}).$$

$\mathbf{T}$ acts on $J_0(N)$ via the action of $\mathbf{T}$ on $S_2(\Gamma_0(N))$. This allows us to view $J_0(N)$ *naturally* as $V/\mathcal{L}$ where V is a complex vector space and $\mathcal{L}$ a lattice in V. The Hecke algebra $\mathbf{T}$ breaks V apart into a direct sum $V = \bigoplus_f V_f$. Fortunately, since the Hecke operators act on V_f and $\mathcal{L}$ in a compatible way, $V_f \cap \mathcal{L}$ is a lattice in V_f. Thus $A_f = V_f/(V_f \cap \mathcal{L})$ is an abelian variety.

Suppose A_f and A_g are not companions, thus f is not congruent to g modulo the actions of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. What is $\#(A_f \cap A_g)$? The general answer isn't known but experimental and theoretical evidence indicates that this number is linked to congruences between f and g.

It is important to emphasize the case when f is a newform of level exactly N in which case $\mathscr{A}_f = E_f$ [[why?]]. Fix such an f, let $E = E_f$ and let

$$d = \dim A_f = \dim_{\mathbf{Q}} E_f = [E_f : \mathbf{Q}].$$

Let ℓ be a prime and consider Tate_ℓ the $\mathbf{Q}_\ell$-adic Tate module of A_f. Recall that

$$\mathrm{Tate}_\ell = \varprojlim{}_\nu A_f[\ell^\nu].$$

Note that $\mathrm{Tate}_\ell \cong \mathbf{Q}_\ell^d$. There is a natural action of $E \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$ on Tate_ℓ. By algebraic number theory

$$E \otimes_{\mathbf{Q}} \mathbf{Q}_\ell = \prod_{\lambda | \ell} E_\lambda$$

where λ runs through primes of E lying over ℓ and E_λ denotes the completion of E with respect to the absolute value induced by λ. Thus Tate_ℓ decomposes as a product

$$\mathrm{Tate}_\ell = \prod_{\lambda | \ell} \mathrm{Tate}_\lambda$$

where Tate_λ is an E_λ vector space. The following lemma will prove useful.

Lemma 12.0.4. *Let the notation be as above. Then for all λ*

$$\dim_{E_\lambda} \mathrm{Tate}_\lambda = 2.$$

Proof. Write $A_f = V_f/\mathcal{L}_f$. The dimension of $\mathcal{L}_f \otimes \mathbf{Q}$ as a vector space over E_f is $\dim V_f/d$ [[UNFINISHED]] $\square$

Now consider Tate_λ which is an E_λ vector space of dimension 2. Since the Hecke operators are defined over $\mathbf{Q}$ it follows that $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts on Tate_ℓ in a way compatible with the action of $E \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$. We thus get a map

$$\rho_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{E\otimes_{\mathbf{Q}}\mathbf{Q}_\ell} \mathrm{Tate}_\ell = \mathrm{GL}_2(E \otimes_{\mathbf{Q}} \mathbf{Q}_\ell) = \prod_\lambda \mathrm{GL}_2(E_\lambda).$$

Thus ρ_ℓ is the direct sum of maps ρ_λ where

$$\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{End}_{E_\lambda}(\mathrm{Tate}_\lambda)$$

gives the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on Tate_λ. If $p \nmid \ell N$ then ρ_λ is unramified at p. In this case it makes sense to consider $\rho_\lambda(\varphi_p)$ where φ_p is a Frobenius element at p. Then $\rho_\lambda(\varphi_p)$ has a well-defined trace and determinent or equivalently a well-defined characteristic polynomial $\Phi(X)$.

Theorem 12.0.5. *Suppose $p \nmid \ell N$. Let $\Phi(X)$ be the characteristic polynomial of $\rho_\lambda(\varphi_p)$. Then*

$$\Phi(X) = X^2 - a_p X + p$$

where a_p is the pth coefficient of the modular form f (thus a_p is the image of T_p in E_f).

Proof. By the Cayley-Hamilton theorem

$$\rho_\lambda(\varphi)^2 - \mathrm{Tr}(\rho_\lambda(\varphi))\rho_\lambda(\varphi) + \det(\rho_\lambda(\varphi)) = 0.$$

Using the Weil pairing one can show that $\mathrm{Tr}(\rho_\lambda(\varphi)) = a_p$. We prove that $\det(\rho_\lambda(\varphi)) = p$. This is a consequence of the fact that

$$\det(\rho_\lambda) = \chi_\ell$$

where χ_ℓ is the ℓth cyclotomic character

$$\chi_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{Z}_\ell^* \subset E_\lambda^*$$

[[Apparently one continues from here by arguing that the Frobenii are dense in some product and so Tchebetarov density implies something.]] □

The simplest case is when A_f is an elliptic curve. In [33] it is shown that $\det(\rho_\ell) = \chi_\ell$. This is proved using the Weil pairing. [[I couldn't find where in silverman, but I didn't look hard.]] More generally we define a Weil pairing

$$\mathrm{Tate}\, A_f \times \mathrm{Tate}\, A_f^\vee \to E_\ell(1).$$

Note that A_f is not necessarily self-dual as was the case when $g = 1$. Note also that the pairing maps into $E_\ell(1)$ instead of $\mathbf{Q}_\ell(1) = \boldsymbol{\mu}_\ell$.

12.1 Computing the Determinent of ρ_λ

Let $f = \sum a_n q^n$ be a newform of weight 2 and level exactly N. Thus f is a common eigenform for the Hecke operators acting on $S_2(\Gamma_0(N))$ and f is orthogonal to the forms of lower weight. Then f gives rise to an abelian subvariety $A = A_f \subset J_0(N)$. There is a canonical action of $\mathbf{T} \otimes \mathbf{Q}$ on A. Since A is the image of $J_0(N)$ under the projection map

π_f, the action of $\mathbf{T} \otimes \mathbf{Q}$ factors through the quotient $E_f = \mathbf{Q}(\dots, a_n, \dots)$. The dimension of A is equal to the degree $[E_f : \mathbf{Q}]$. The module

$$\mathrm{Tate}_\ell(A) = \varprojlim A[\ell^\nu]$$

is free of rank 2 over

$$E_f \otimes \mathbf{Q}_\ell = \prod_{\lambda|\ell} E_\lambda.$$

Thus we obtain a decomposition

$$\mathrm{Tate}_\ell(A) = \prod_{\lambda|\ell} \mathrm{Tate}_\lambda(A).$$

For each λ this gives a map

$$\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{E_\lambda} \mathrm{Tate}_\lambda = \mathrm{GL}(2, E_\lambda).$$

If $p \nmid \ell N$ we consider $\rho_\lambda(\mathrm{Frob}_p)$ where $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is a Frobenius element for p (note the change in notation, before we called the Frobenius element φ_p). We next discuss why $\rho_\lambda(\mathrm{Frob}_p)$ satisfies $X^2 - a_p X + p = 0$.

Theorem 12.1.1. *Let the notation be as above. Then* $\mathrm{Tr}(\rho_\lambda(\mathrm{Frob}_p)) = a_p$ *and* $\det(\rho_\lambda(\mathrm{Frob}_p)) = p$.

The main step is to show that $\det \rho_\lambda = \chi_\ell$. In order to do this we must first introduce the Weil pairing on $\mathrm{Tate}_\ell A$. We need a little background on abelian varieties.

12.2 Duality and Polarizations

Let A be an abelian variety. A **divisor** on A is an integral linear combination $\sum m_i D_i$ of codimension 1 algebraic subvarieties D_i. If f is a rational function on A then we define the **divisor of** f to be

$$(f) = \sum (\mathrm{ord}_D f) D.$$

(See Hartshorne [7] II.6 for the definition of $\mathrm{ord}_D f$.)

Two divisors D_1 and D_2 are **linearly equivalent** (denoted $D_1 \sim D_2$) if $D_1 - D_2 = (f)$ for some rational function f. We denote by $\mathrm{Pic}(A)$ the group of divisors modulo linear equivalence. Two divisors D_1 and D_2 are **algebraically equivalent** if D_1 is linearly equivalent to a translate of D_2 (i.e. $D_1 \sim D_2 + a$ for some $a \in A$). The subgroup of $\mathrm{Pic}(A)$ of divisors algebraically equivalent to zero is denoted $\mathrm{Pic}^0(A)$. It is a fact that $\mathrm{Pic}^0(A)$ is an abelian variety called the **dual** of A and denoted $A^\vee$. In the more general arithmetic setting $\mathrm{Pic}^0(A)$ is replaced by isomorphism classes of certain invertible sheaves of degree 0. A **polarization** of A is an isogeny $A \to \mathrm{Pic}^0(A)$ of the form

$$a \mapsto t_a^* \mathcal{L} \otimes \mathcal{L}^{-1}$$

where t_a^* is translation by A and $\mathcal{L}$ is an ample invertible sheaf. For further details sees Milne's article [16] or Murty's book [19]. [[I've given two different definitions. Thus I've corned myself and should probably rewrite all of the above as soon as I have a better idea of what is going on from Milne's point of view.]]

Since $A \subset J_0(N)$, autoduality gives an isomorphism $J_0(N) \xrightarrow{\sim} J_0(N)^\vee$ so we obtain a canonical polarization $\varphi : A \to A^\vee$. The diagram is

$$\begin{array}{ccc} J_0(N) & \xrightarrow{\textit{autoduality}} & J_0(N)^\vee \\ \cup & & \downarrow \\ A & \xrightarrow{\textit{polarization}} & A^\vee \end{array}$$

12.3 The Weil Pairing

Let $A = A_f$ be the abelian variety connected to the modular form f. One can define a bilinear pairing

$$A[n] \times A^\vee[n] \to \boldsymbol{\mu}_n$$

which is a natural generalization of the Weil pairing for an elliptic curve. See §16 of Milne's arcticle [16]. The canonical polarization of A is an isogeny $\varphi : A \to A^\vee$. Combining this with the pairing gives a map

$$A[n] \times A[n] \to \boldsymbol{\mu}_n$$

which is alternating. But it is not a perfect pairing. Fortunately, upon taking limits, we obtain a perfect pairing

$$(\cdot,\cdot) : \mathrm{Tate}_\ell(A) \times \mathrm{Tate}_\ell(A) \to \mathbf{Q}_\ell(1) = \varprojlim \boldsymbol{\mu}_{\ell^\nu} \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell.$$

If $t \in \mathbf{T}_0 \otimes \mathbf{Q}_\ell$ then $(ta, a') = (a, ta')$.

Using the polarization φ we can construct a map $\mathrm{End}(A) \to \mathrm{End}(A) \otimes \mathbf{Q}$ by

$$t \mapsto t^* = \varphi^{-1} t^\vee \varphi.$$

The map $t \mapsto t^*$ is called the **Rosati involution** and has the property that $(ta, a') = (a, t^*a')$ for $a, a' \in \mathrm{Tate}_\ell(A)$.

Now restrict the form

$$(\cdot,\cdot) : \mathrm{Tate}_\ell(A) \times \mathrm{Tate}_\ell(A) \to \mathbf{Q}_\ell(1)$$

to $\mathrm{Tate}_\lambda(A)$ obtaining

$$\langle\cdot,\cdot\rangle : \mathrm{Tate}_\lambda(A) \times \mathrm{Tate}_\lambda(A) \to \mathbf{Q}_\ell(1).$$

The form $\langle\cdot,\cdot\rangle$ is alternating and nondegenerate. The nondegeneracy follows from the nondegeneracy of $(\cdot,\cdot)$ and the fact that $(\mathrm{Tate}_\lambda, \mathrm{Tate}_{\lambda'}) = 0$ when $\lambda \neq \lambda'$. The pairing $\langle\cdot,\cdot\rangle$ is Galois equivariant in the following sense. If $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ and $x, y \in \mathrm{Tate}_\lambda$ then

$$\langle\sigma x, \sigma y\rangle = \sigma\langle x, y\rangle = \chi_\ell(\sigma)\langle x, y\rangle.$$

Note that σ acts on $\mathbf{Q}_\ell(1)$ by multiplication by $\chi_\ell(\sigma)$.

12.4 The Fancy Proof

There are two proofs of the theorem, a fancy proof and a concrete proof. We first present the fancy proof. The pairing $\langle\cdot,\cdot\rangle$ is a alternating and bilinear form so it defines a map

$$\langle\cdot,\cdot\rangle : \Lambda^2_{E_\lambda} \mathrm{Tate}_\lambda \to \mathbf{Q}_\ell(1).$$

It is not *a priori* true that we can take the wedge product over E_λ instead of $\mathbf{Q}_\ell$, but we can because $\langle tx, y\rangle = \langle x, ty\rangle$ for any $t \in E_\lambda$. Let $D = \Lambda^2 \operatorname{Tate}_\lambda$ and note that $\dim_{E_\lambda} D = 1$.

There is a trace map which gives an identification

$$\operatorname{Hom}_{\mathbf{Q}_\ell}(D, \mathbf{Q}_\ell(1)) \cong \operatorname{Hom}_{E_\lambda}(D, E_\lambda(1))$$

Thus $\langle \cdot, \cdot \rangle$ is an isomorphism $D \cong E_\lambda(1)$. [[I don't see this implication right now.]] It now follows that $\det \rho_\lambda = \chi_\ell$.

12.5 The Concrete Proof

Now we consider the concrete proof. Suppose $\sigma \in \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, then we must show that $\det(\sigma) = \chi_\ell(\sigma)$. Choose vectors $x, y \in \operatorname{Tate}_\lambda$ which are linearly independent over E_λ. Let $a, b, c, d \in E_\lambda$ be defined by $\sigma(x) = ax + cy$ and $\sigma(y) = bx + dy$. Let $\delta = ad - bc - \chi_\ell(\sigma)$, thus we want to show that $\delta = 0$. We have

$$\begin{aligned}
\chi_\ell(\sigma)\langle x, y\rangle &= \langle \sigma x, \sigma y\rangle \\
&= \langle ax + cy, bx + dy\rangle \\
&= \langle ax, bx\rangle + \langle ax, dy\rangle + \langle cy, bx\rangle + \langle cy, dy\rangle \\
&= \langle ax, dy\rangle + \langle cy, bx\rangle \\
&= \langle adx, y\rangle - \langle bcx, y\rangle = \langle (ad - bc)x, y\rangle
\end{aligned}$$

To see that $\langle ax, bx\rangle = 0$ we noted that

$$\langle ax, bx\rangle = \langle abx, x\rangle = -\langle x, abx\rangle = -\langle ax, bx\rangle.$$

[[Of course it looks like the next step is too use nondegeneracy to conclude that $\chi_\ell(\sigma) = ad - bc$, but it is not clear to me how to safely do this. My problem is that the expression $(ad - bc)\langle x, y\rangle$ does not really make sense unless we know that $ad - bc \in \mathbf{Q}_\lambda(1)$ and, even then, why is this equal to the last expression? I clearly lack some crucial bit of knowledge about $\mathbf{Q}_\lambda(1)$]]

12.6 The Construction for $X_1(N)$

Let $f = \sum a_n q^n$ be a newform for $\Gamma_1(N)$, so $f|T_n = a_n f$ for all $n \geq 1$ and $f|\langle d\rangle = \varepsilon(d) f$ for all $d \in (\mathbf{Z}/N\mathbf{Z})^*$. In the case of $X_1(N)$, E_f is either totally real or it can be a CM field. If it is a CM field then it has a canonical complex conjugation $\bar{\cdot} : a_n \mapsto \varepsilon(n)^{-1} a_n$. The analogous formulas are

$$\begin{aligned}
\operatorname{Tr}(\rho_\ell(\operatorname{Frob}_p)) &= a_p \\
\det(\rho_\ell(\operatorname{Frob}_p)) &= \varepsilon(p)p
\end{aligned}$$

There are some details concerning signs which we must consider. First, for every p we can define T_p as a correspondence on $X_1(N)$ with α and β defined in an analogous way to the way they were defined in the case of $X_0(N)$. But now we must choose if the corresponding endomorphism of $J_1(N)$ should be $(T_p)_*$ or $(T_p)^*$. The second choice concerns what happens in the Eichler-Shimura relation. Recall that $X_1(N)$ classifies pairs (E, p) where E is an elliptic curve and p is a point of order N. Equivalently, $X_1(N)$ classifies pairs consisting of an elliptic curve and an embedding $\mathbf{Z}/N\mathbf{Z} \hookrightarrow E$. But this is a bad way to look at this.

A better way is to view $X_1(N)$ as classifying pairs consisting of an elliptic curve and an embedding $\boldsymbol{\mu}_N \hookrightarrow E$ because the associated Tate curve has a copy of $\boldsymbol{\mu}_N$ naturally sitting in it. When passing to $J_1(N)$ we take the Picard functorality such that if $T_p : J_1(N) \to J_1(N)$ then $(T_p^\vee)^*$ on $H^0(X_1(N), \Omega^1) = S_2(\Gamma_1(N))$ is the classical T_p.

Chapter 13

The Gorenstein Property

Suppose $f = \sum a_n q^n$ is a newform of exact level N and weight 2 for the congruence subgroup $\Gamma_0(N)$. Let $E = \mathbf{Q}(\ldots, a_n, \ldots)$ and let λ be a place of E lieing over the prime ℓ of $\mathbf{Q}$. The action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the ℓ-adic Tate module of the abelian variety A_f associated to f gives rise to a representation

$$\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(E_\lambda)$$

with $\det \rho_\lambda = \chi_\ell$ and $\mathrm{Tr}\,\rho_\lambda(\mathrm{Frob}_p) = a_p$ for $p \not| \ell N$. We ask, how can we reduce this representation module λ?

Lemma 13.0.1. *Let $\mathcal{O}$ be the ring of integers of E_λ. Then ρ_λ is equivalent to a representation which takes values in $\mathrm{GL}_2(\mathcal{O})$.*

Proof. View $\mathrm{GL}_2(E_\lambda)$ as the group of automorphisms of a 2 dimensional E_λ-vector space V. A lattice is a free $\mathcal{O}$-module of rank 2 such that $L \otimes E_\lambda \cong V$. It suffices to find an $\mathcal{O}$-lattice L in V which is invariant under the action (via ρ_λ) of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. For then the matrices of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ with respect to a basis of V consisting of vectors from L will have coefficients in $\mathcal{O}$. Choose any lattice $L_0 \subset V$. Since L_0 is discrete and $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is compact the set of lattices gL_0 with $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is finite. Let $L = \sum gL_0$ be the sum of the finitely many conjugates of L_0, then L is Galois invariant. $\square$

It now makes sense to let $\overline{\rho}_\lambda = \rho_\lambda \mod \lambda$. This is expressed by the diagram

$$\begin{array}{ccc} \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) & \xrightarrow{\rho_\lambda} & \mathrm{GL}_2(\mathcal{O}) \\ & \searrow & \downarrow \\ & & \mathrm{GL}_2(\mathcal{O}/\lambda) \end{array}$$

The drawback to this approach is that $\overline{\rho}_\lambda$ is not intrinsic because the definition depends on the choice of L_0. In fact, it is possible to obtain nonisomorphic representations $\overline{\rho}_\lambda$ by different choices of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathcal{O})$. By nonisomorphic we mean nonisomorphic over $\mathcal{O}$ (the representations are isomorphic over E_λ).

The good definition is to let $\overline{\rho}_\lambda$ be the semisimplification of the reduction of ρ_λ modulo λ. By the semisimplification of a representation we mean the direct sum of the Jordan-Hölder factors in a filtration of V. Suppose $\sigma : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{F})$ is a mod ℓ Galois representation. If $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ then the characteristic polynomial of the action of g in the mod λ representation is the same as the characteristic polynomial of g in the semisimplification. The representation $\overline{\rho}_\lambda$ is semisimple (since it is a semisimplification) and $\det \overline{\rho}_\lambda = \chi_\ell \mod \ell$ is the mod ℓ cyclotomic character, so for example $\chi_\ell(\mathrm{Frob}_p) = p \mod \ell$. Furthermore,

$\mathrm{Tr}(\overline{\rho}_\lambda(\mathrm{Frob}_p)) = a_p \mod \lambda$. This shows that the characteristic polynomials corresponding to $\overline{\rho}_\lambda$ are independent of our choice of an isomorphic representation with values in $\mathrm{GL}_2(\mathcal{O})$. Thus the following theorem implies that $\overline{\rho}_\lambda$ only depends on f and λ.

Theorem 13.0.2 (Brauer-Nesbit). *Suppose $\rho_1, \rho_2 : G \to \mathrm{GL}(V)$ are two finite dimensional semisimple representations of a group G over a finite field k. Assume furthermore that for every $g \in G$ the characteristic polynomial of $\rho_1(g)$ is the same as the characteristic polynomial of $\rho_2(g)$. Then ρ_1 and ρ_2 are equivalent.*

The proof can be found in section 30.16 of [3].

The Hecke operators T_n act on the space $S_2(\Gamma_0(N))$. Let $\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(S_2(\Gamma_0(N)))$ and recall that $\mathbf{T}$ is a commutative ring which is a $\mathbf{Z}$-module of rank equal to $\dim_{\mathbf{C}} S_2(\Gamma_0(N))$. Let $\mathbf{m}$ be a maximal ideal of $\mathbf{T}$ and let $k = \mathbf{T}/\mathbf{m} = \mathbf{F}_{\ell^\nu}$ be the residue field.

Proposition 13.0.3. *There is a (unique) semisimple representation*

$$\rho_{\mathbf{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(k)$$

such that $\rho_{\mathbf{m}}$ is unramified outside ℓN and

$$\mathrm{Tr}(\rho_{\mathbf{m}}(\mathrm{Frob}_p)) = T_p \mod \mathbf{m}$$
$$\det(\rho_{\mathbf{m}}(\mathrm{Frob}_p)) = p \mod \mathbf{m}$$

Proof. It is enough to prove the assertion with $\mathbf{T}$ replaced by

$$\mathbf{T}_0 = \mathbf{Z}[\{\ldots, T_n, \ldots : (n, N) = 1\}].$$

Indeed, a maximal ideal of $\mathbf{T}$ pulls back to a maximal ideal of $\mathbf{T}_0$ [[why is this true in this case?]] and $k_0 = \mathbf{T}_0/\mathbf{m}_0 \subset k$. Now

$$\mathbf{T}_0 \subset \mathbf{T}_0 \otimes \mathbf{Q} = \prod_{i=1}^{t} E_i$$

with the E_i number fields. Let $\mathcal{O}_{E_i}$ be the ring of integers of E_i and let $\mathcal{O} = \prod \mathcal{O}_{E_i}$. Then

$$\mathbf{m}_0 \subset \mathbf{T}_0 \subset \prod \mathcal{O}_{E_i}.$$

By the going-up theorem there is a maximal ideal $\lambda \subset \prod \mathcal{O}_{E_i}$ lieing over $\mathbf{m}_0$, and $k_0 \subset \mathcal{O}/\lambda$. Using the construction outlined above (the field E is the only E_i for which $E_i \cap \lambda \neq 0$) we can make a representation

$$\overline{\rho}_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathcal{O}/\lambda)$$

with $\det \overline{\rho}_\lambda = \chi_\ell$ and $\mathrm{Tr}\,\overline{\rho}_\lambda = T_p \mod \mathbf{m}$ (because of the way we have set things up T_p plays the role of a_p). Thus this representation has the required properties, but it takes values in $\mathrm{GL}_2(\mathcal{O}/\lambda)$ instead of k_0. Since the characteristic polynomial of every $\overline{\rho}_\lambda(g)$ for $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ has coefficients in the subfield $k_0 \subset \mathcal{O}/\lambda$ there is a model for ρ_λ over k_0. This is the extra bit of information we need to complete the proof. It is a fact about finite fields and is somehow related to the fact that the Brauer group of a finite field is trivial. [[It would be *great* to have a reference for this fact about finite fields because, judging from the audience, it is not an extremely well-known result.]] □

Let us look at this construction in another way. Write $\mathbf{T}_0 \otimes \mathbf{Q} = E_1 \times \cdots \times E_t$ and recall that E_i corresponds to a newform of level $M|N$ (one can obtain E_i by adjoining the coefficients of some newform of level $M|N$ to $\mathbf{Q}$). Also, $J_0(N)$ is isogenous to a product $A_1 \times \cdots \times A_t$. Consider one of the factors, say E_1, and suppose, to fix ideas, that it corresponds to a newform of exact level N. Since $\mathrm{Tate}_\ell A_1$ is free of rank 2 over $E_1 \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$, we obtain a 2-dimensional representation ρ_λ. Reducing mod λ and semisimplifying gives the representation constructed in the above proposition. But it is also possible that one of the fields E_i corresponds to a newform f of level M properly dividing N. Then the associated abelian variety A_i has too large of dimension. So [[somehow?!]] we repeat the whole construction with $J_0(N)$ to get a 2 dimensional representation.

Let $A = A_1$ be as above. Then $\mathbf{T}$ acts on A_1 which we view as an abelian subvariety of $J_0(N)$. Let $\overline{\mathbf{T}}$ be the quotient of $\mathbf{T}$ by the kernel of the map $\mathbf{T} \to \mathrm{End}\, A_1$. Although $\overline{\mathbf{T}}$ sits in $\mathcal{O}_1$ which is the ring of integers of a field, $\overline{\mathbf{T}}$ might not be integrally closed in $\mathcal{O}_1$. Consider the usual $\mathbf{Z}_\ell$-adic Tate module $\mathrm{Tate}_\ell(A_1) \cong \mathbf{Z}_\ell^{2\dim A}$ ($\dim A = [E_1 : \mathbf{Q}]$). This Tate module has a faithful action of $\overline{\mathbf{T}} \otimes \mathbf{Z}_\ell$. By the theory of semilocal rings this ring decomposes as a product

$$\overline{\mathbf{T}} \otimes \mathbf{Z}_\ell = \prod_{\mathbf{m}|\ell} \overline{\mathbf{T}}_{\mathbf{m}}.$$

Thus Tate_ℓ decomposes as a product $\prod_{\mathbf{m}} \mathrm{Tate}_{\mathbf{m}}(A_1)$. It would be nice if Tate_ℓ were free of rank 2 over $\overline{\mathbf{T}} \otimes \mathbf{Z}_\ell$ but this is not known to be true in general, although it has often been verified in special cases. For this to be true we must have that, for all $\mathbf{m}|\ell$, $\mathrm{Tate}_{\mathbf{m}}$ is free of rank 2 over $\overline{\mathbf{T}}_{\mathbf{m}}$.

Next we put things in a finite context instead of a projective limit context. Let $J = J_0(N)$, then by Albanese or Picard functoriality $\mathbf{T} \subset \mathrm{End}\, J$. Let $\mathbf{m} \subset \mathbf{T}$ be a maximal ideal. Let

$$J[\mathbf{m}] = \{t \in J(\overline{\mathbf{Q}}) : xt = 0 \text{ for all } x \in \mathbf{m}\}.$$

Note that $J[\mathbf{m}] \subset J[\ell]$ where ℓ is the rational prime lying in $\mathbf{m}$. Now $J[\ell]$ is an $\mathbf{F}_\ell$ vector space of rank $2g$ where g is the genus of $X_0(N)$. Although it is true that $J[\ell]$ is a $\mathbf{T}/\ell$-module it is not convenient to work with $\mathbf{T}/\ell$ since it might not be a product of field and there can be some unpleasant ramification. It is more convenient to work with $J[\mathbf{m}]$ since $\mathbf{T}/\mathbf{m}$ is a field. Thus, via another optic, $J[\mathbf{m}]$ is a $k[G]$-module where $k = \mathbf{T}/\mathbf{m}$ and $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. The naive hope is that $J[\mathbf{m}]$ is a model for $\rho_{\mathbf{m}}$, at least when $\rho_{\mathbf{m}}$ is irreducible. This does not quite work, but we do have the following

Theorem 13.0.4. *$J[\mathbf{m}]$ is a model for $\rho_{\mathbf{m}}$ when $\ell \neq 2$ and $\ell \nmid N$.*

13.1 The Gorenstein Property

Let

$$\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}\, J_0(N)$$

and let $\mathbf{m} \subset \mathbf{T}$ be a maximal ideal. Let $\ell \in \mathbf{m}$ be the prime of $\mathbf{Z}$ over which $\mathbf{m}$ lies. We have constructed a representation

$$\rho_{\mathbf{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{T}/\mathbf{m})$$

which is semisimple by construction. It is unramified outside ℓN and for any prime $p \nmid \ell N$,

$$\mathrm{Tr}(\rho_{\mathbf{m}}(\mathrm{Frob}_p)) = T_p \mod \mathbf{m}$$
$$\det(\rho_{\mathbf{m}}(\mathrm{Frob}_p)) = p \mod \mathbf{m}$$

We will usually be interested in the case when $\rho_{\mathbf{m}}$ is irreducible. Let $\mathbf{T}_{\mathbf{m}} = \varprojlim \mathbf{T}/\mathbf{m}^i\mathbf{T}$ denote the completion of $\mathbf{T}$ at $\mathbf{m}$. Note that $\mathbf{T} \otimes_{\mathbf{Z}} \mathbf{Z}_\ell = \prod_{\mathbf{m}|\ell} \mathbf{T}_{\mathbf{m}}$. Our goal is to prove that if $\mathbf{m} \not| 2N$ and $\rho_{\mathbf{m}}$ is irreducible, then $\mathbf{T}_{\mathbf{m}}$ is Gorenstein.

Definition 13.1.1. Let $\mathcal{O}$ be a complete discrete valuation ring. Let T be a local $\mathcal{O}$-algebra which as a module if finite and free over $\mathcal{O}$. Then T is a **Gorenstein $\mathcal{O}$-algebra** if there is an isomorphism of T-modules $T \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}}(T, \mathcal{O})$.

Thus $\mathbf{T}_m$ is Gorenstein if $\mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{T}_{\mathbf{m}}, \mathbf{Z}_\ell) \xrightarrow{\sim} \mathbf{T}_{\mathbf{m}}.$ as $\mathbf{T}_{\mathbf{m}}$ modules. Intuitively this means $\mathbf{T}_{\mathbf{m}}$ is "autodual".

Theorem 13.1.2. *Let $J = J_0(N)$, let the representation $\rho_{\mathbf{m}}$ be the as constructed in Proposition ?? and assume $\rho_{\mathbf{m}}$ is irreducible and $\mathbf{m} \not| 2N$. Let*

$$J[\mathbf{m}] = \{a : xa = 0 \text{ for all } x \in \mathbf{m}\} \subset J[\ell].$$

Then $\dim_{\mathbf{T}/\mathbf{m}} J[\mathbf{m}] = 2$ and $J[\mathbf{m}]$ as a Galois module is a 2-dimensional representation giving rise to $\rho_{\mathbf{m}}$.

[[Does giving rise mean the char poly's are all the same or that the semisimplifications are isomorphic?]] An easy argument shows that the theorem implies $\mathbf{T}_{\mathbf{m}}$ is Gorenstein.

We first consider the structure of $W = J[\mathbf{m}]$. Suppose the two dimensional representation

$$\rho_{\mathbf{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{\mathbf{T}/\mathbf{m}} V$$

constructed before is irreducible. Consider the semisimplification $W^{\text{s.s.}}$ of W, thus $W^{\text{s.s.}}$ is the direct sum of its Jordan-Hölder factors as a $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-module. Mazur proved the following theorem.

Theorem 13.1.3. *There is some integer $t \geq 0$ so that*

$$W^{s.s.} \cong V \times \cdots \times V = V^t.$$

If $\rho_{\mathbf{m}}$ is in fact *absolutely irreducible* then it is a result of Boston, Lenstra, and Ribet [2] that $W \cong V \times \cdots \times V$. A representation is absolutely irreducible if it is irreducible over the algebraic closure. It can be shown that if $\ell \neq 2$ and $\rho_{\mathbf{m}}$ is irreducible then $\rho_{\mathbf{m}}$ must be absolutely irreducible.

The construction of W is nice and gentle whereas the construction of V is accomplished via brute force.

Proof. (Mazur) We want to compare V with W. Let $d = \dim W$. Let

$$W^* = \mathrm{Hom}_{\mathbf{T}/\mathbf{m}}(W, \mathbf{T}/\mathbf{m}(1))$$

where $\mathbf{T}/\mathbf{m}(1) = \mathbf{T}/\mathbf{m} \otimes_{\mathbf{Z}} \boldsymbol{\mu}_\ell$. We need to show that

$$W^{\text{s.s.}} \oplus W^{*\text{s.s.}} \cong V \times \cdots \times V = V^d$$

as representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Note that each side is a semisimple module of dimension $2d$. To obtain the isomorphism we show that the two representations have the same characteristic polynomials so they are isomorphic.

We want to show that the characteristic polynomial of Frob_p is the same for both $W^{\text{s.s.}} \oplus W^{\text{s.s.}}$ and $V \times \cdots \times V$. The characteristic polynomial of Frob_p on V is $X^2 - T_pX + p = (X-r)(X-pr^{-1})$ where r lies in a suitable algebraic closure. It follows that the characteristic polynomial of Frob_p on V^d is $(x-r)^d(x-pr^{-1})^d$. On W the characteristic polynomial of Frob_p is $(X-\alpha_1)\cdots(X-\alpha_d)$ where α_i is either r or pr^{-1}. This is because Eichler-Shimura implies Frob_p must satisfy $\mathrm{Frob}_p^2 - T_p\,\mathrm{Frob}_p + p = 0$. [[I don't see this implication.]] On W^* the characteristic polynomial of Frob_p is $(X-p\alpha_1^{-1})\cdots(X-p\alpha_d^{-1})$. [[This is somehow tied up with the definition of W^* and I can't quite understand it.]] Thus on $W\oplus W^*$, the characteristic polynomial of Frob_p is

$$\prod_{i=1}^{d}(X-\alpha_i)(X-p\alpha_i^{-1}) = \prod_{i=1}^{d}(X-r)(X-pr^{-1}) = (X-r)^d(X-pr^{-1})^d.$$

Therefore the characteristic polynomial of Frob_p on $W\oplus W^*$ is the same as the characteristic polynomial of Frob_p on $V\times\cdots\times V$. The point is that although the α_i could all *a priori* be r or pr^{-1}, by adding in W^* everything pairs off correctly. [[I don't understand why we only have to check that the two representations agree on Frob_p. There are lots of other elements in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, right?]] □

We next show that $J[\mathbf{m}]\neq 0$. This does not follow from the theorem proved above because it does not rule out the possibility that $t=0$ and hence $W^{\text{s.s.}}\cong 0$. Suppose $J[\mathbf{m}]=0$, then we will show that $J[\mathbf{m}^i]=0$ for all $i\geq 1$. We consider the ℓ-divisible group $J_{\mathbf{m}}=\cup_i J[\mathbf{m}^i]$. To get a better feel for what is going on, temporarily forget about $\mathbf{m}$ and just consider the Tate module corresponding to ℓ.

It is standard to consider the Tate module

$$\mathrm{Tate}_\ell\, J = \varprojlim J[\ell^i] \cong \mathbf{Z}_\ell^{2\dim J}.$$

It is completely equivalent to consider

$$J_\ell := \cup_{i=1}^{\infty} J[\ell^i] \xrightarrow{\sim} (\mathbf{Q}_\ell/\mathbf{Z}_\ell)^{2\dim J}.$$

Note that since $\mathbf{Q}_\ell/\mathbf{Z}_\ell$ is not a ring the last isomorphism must be viewed as an isomorphism of abelian groups. In [14] Mazur called $\mathrm{Tate}_\ell\, J\cong \mathrm{Hom}(\mathbf{Q}_\ell/\mathbf{Z}_\ell, J_\ell)$ the covariant Tate module. Call

$$\mathrm{Tate}_\ell^*\, J := \mathrm{Hom}(J_\ell, \mathbf{Q}_\ell/\mathbf{Z}_\ell) \cong \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\ell\, J, \mathbf{Z}_\ell)$$

the contravariant Tate module. [[Why are the last two isomorphic?]] The covariant and contravariant Tate modules are related by a Weil pairing $J[\ell^i]\times J[\ell^i]\to \boldsymbol{\mu}_{\ell^i}$. Taking projective limits we obtain a pairing

$$\langle\cdot,\cdot\rangle : \mathrm{Tate}_\ell\, J\times \mathrm{Tate}_\ell\, J \to \mathbf{Z}_\ell(1) = \varprojlim \boldsymbol{\mu}_{\ell^i}.$$

This gives a map

$$\mathrm{Tate}_\ell\, J \to \mathrm{Hom}(\mathrm{Tate}_\ell\, J, \mathbf{Z}_\ell(1)) = (\mathrm{Tate}_\ell^*\, J)(1)$$

where $(\mathrm{Tate}_\ell^*\, J)(1) = (\mathrm{Tate}_\ell^*\, J)\otimes \mathbf{Z}_\ell(1)$. $\mathbf{Z}_\ell(1)$ is a $\mathbf{Z}_\ell$-module where

$$\sum a_i\ell^i\cdot\zeta = \zeta^{\sum a_i\ell^i}$$

[[This should probably be said long ago.]] This pairing is not a pairing of $\mathbf{T}$-modules, since if $t\in\mathbf{T}$ then $\langle tx,y\rangle = \langle x, t^\vee y\rangle$. It is more convenient to use an adapted pairing defined as follows. Let $w=w_V\in \mathrm{End}\, J_0(N)$ be the Atkin-Lehner involution so that $t^\vee = wtw$. Define a new $\mathbf{T}$-compatible pairing by $[x,y] := \langle x, wy\rangle$. Then

$$[tx,y] = \langle tx, wy\rangle = \langle x, t^\vee wy\rangle = \langle x, wty\rangle = [x,ty].$$

The pairing $[\cdot,\cdot]$ defines an isomorphism of $\mathbf{T}\otimes\mathbf{Z}_\ell$-modules

$$\mathrm{Tate}_\ell J \xrightarrow{\sim} \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\ell J, \mathbf{Z}_\ell(1)).$$

Since $\mathbf{Z}_\ell(1)$ is a free module of rank 1 over $\mathbf{Z}_\ell$ a suitable choice of basis gives an isomorphism of $\mathbf{T}\otimes\mathbf{Z}_\ell$-modules

$$(\mathrm{Tate}_\ell^* J)(1) \cong \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\ell J, \mathbf{Z}_\ell(1)) \cong \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\ell J, \mathbf{Z}_\ell) = \mathrm{Tate}_\ell^* J.$$

Thus $\mathrm{Tate}_\ell J \xrightarrow{\sim} \mathrm{Tate}_\ell^* J$.

Proof. (That $J[\mathbf{m}] \neq 0$.) The point is that the contravariant Tate module $\mathrm{Hom}(J_\ell, \mathbf{Q}_\ell/\mathbf{Z}_\ell)$ is the Pontrjagin dual of T_ℓ. How does this relate to $\mathrm{Tate}_{\mathbf{m}} J$? Since $\mathbf{T}\otimes\mathbf{Z}_\ell \cong \prod_{\mathbf{m}|\ell} \mathbf{T}_{\mathbf{m}}$, $\mathrm{Tate}_\ell J \cong \prod_{\mathbf{m}|\ell} \mathrm{Tate}_{\mathbf{m}} J$ so we can define $\mathrm{Tate}_{\mathbf{m}}^* J := \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_{\mathbf{m}} J, \mathbf{Z}_\ell)$. Weil proved that $\mathrm{Tate}_{\mathbf{m}} J \cong Tate_{\mathbf{m}}^* J$ is nonzero. View $\mathrm{Tate}_{\mathbf{m}}^* J$ as being dual to $J_{\mathbf{m}}$ in the sense of Pontrjagin duality and so $(\mathrm{Tate}_{\mathbf{m}}^* J)/(\mathbf{m}\,\mathrm{Tate}_{\mathbf{m}}^* J)$ is dual to $J[\mathbf{m}]$. If $J[\mathbf{m}] = 0$ then this quotient is 0, so Nakayama's lemma would imply that $\mathrm{Tate}_{\mathbf{m}}^* J = 0$. This would contradicts Weil's assertion. Therefore $J[\mathbf{m}] \neq 0$. $\square$

13.2 Proof the Gorenstein Property

We are considering the situation with respect to $J_0(N)$ although we could consider $J_1(N)$. Let $\mathbf{T} \subset \mathrm{End}\, J_0(N)$ be the Hecke algebra and let $\mathbf{m} \subset \mathbf{T}$ be a maximal ideal. Let ℓ be the characteristic of the residue class field $\mathbf{T}/\mathbf{m}$. Let $\mathbf{T}_{\mathbf{m}} = \varprojlim \mathbf{T}/\mathbf{m}^i\mathbf{T}$. Then $\mathbf{T}\otimes_{\mathbf{Z}}\mathbf{Z}_\ell = \prod_{\mathbf{m}|\ell} \mathbf{T}_{\mathbf{m}}$. [[I want to put a good reference for this Atiyah-Macdonald like fact here.]] Each $\mathbf{T}_{\mathbf{m}}$ acts on $\mathrm{Tate}_\ell J_0(N)$ so we obtain a product decomposition

$$\mathrm{Tate}_\ell J_0(N) = \prod_{\mathbf{m}|\ell} \mathrm{Tate}_{\mathbf{m}} J_0(N).$$

We have the following two facts:

1. $\mathrm{Tate}_{\mathbf{m}} J_0(N) \neq 0$
2. $\mathrm{Tate}_\ell J_0(N)$ is $\mathbf{T}\otimes\mathbf{Z}_\ell$-autodual and each $\mathrm{Tate}_{\mathbf{m}} J_0(N)$ is $\mathbf{Z}_\ell$-autodual.

[[autoduality for which dual? I think it is the linear dual since this is used.]]

Let $W = J[\mathbf{m}]$, then the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on W gives a representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ over the field $\mathbf{T}/\mathbf{m}$. We compared W with a certain two dimensional representation $\rho_{\mathbf{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to V$ over $\mathbf{T}/\mathbf{m}$. Assume unless otherwise stated that V is irreducible as a $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-module. Let $\mathrm{Tate}_\ell = \mathrm{Tate}_\ell J_0(N)$ and $\mathrm{Tate}_{\mathbf{m}} = \mathrm{Tate}_{\mathbf{m}} J_0(N)$. A formal argument due to Mazur showed that

$$W^{\text{s.s.}} \cong V \times \cdots \times V = V^{\oplus t}.$$

We have not yet determined t but we would like to show that $t = 1$.

Definition 13.2.1. The **Pontrjagin dual** of a module M is the module $M^\wedge := \mathrm{Hom}(M, \mathbf{Q}/\mathbf{Z})$ where M is viewed as an abelian group (if M is topological, only take those homomorphisms whose kernel is compact). The **linear dual** of a module M over a ring R is the module $M^* = \mathrm{Hom}_R(M, R)$.

Exercise 13.2.2. Note that $(\mathbf{Q}_\ell/\mathbf{Z}_\ell)^\wedge = \mathbf{Z}_\ell$ and $\mathbf{Z}_\ell^\wedge = \mathbf{Q}_\ell/\mathbf{Z}_\ell$.

Solution. We can think of $\mathbf{Q}_\ell/\mathbf{Z}_\ell$ as

$$\{\sum_{n=-k}^{-1} a_n \ell^n : k > 0 \text{ and } 0 \le a_i < \ell\}.$$

Let $(b_i) \in \mathbf{Z}_\ell$ so $b_i \in \mathbf{Z}/\ell^i\mathbf{Z}$ and $b_{i+1} \equiv b_i \pmod{\ell^i}$. Define a map $\mathbf{Q}_\ell/\mathbf{Z}_\ell \to \mathbf{Q}/\mathbf{Z}$ by $1/\ell^i \mapsto b_i/\ell^i$. To check that this is well-defined it suffices to check that $1/\ell^i$ maps to the same place as $\ell \cdot 1/\ell^{i+1}$. Now $1/\ell^i \mapsto b_i/\ell^i$ and

$$\ell \cdot 1/\ell^{i+1} \mapsto \ell \cdot b_{i+1}/\ell^{i+1} = b_{i+1}/\ell^i.$$

So we just need to check that

$$b_{i+1}/\ell^i \equiv b_i/\ell^i \pmod{\mathbf{Z}}.$$

This is just the assertion that $(b_{i+1} - b_i)/\ell^i \in \mathbf{Z}$ which is true since $b_{i+1} \equiv b_i \pmod{\ell^i}$. $\square$

Proposition 13.2.3. *Let the notation be as above, then $t > 0$.*

Proof. The idea is to use Nakayama's lemma to show that if $t = 0$ and hence $W = 0$ then $\mathrm{Tate}_{\mathbf{m}} = 0$ which is clearly false. But the relation between W and $\mathrm{Tate}_{\mathbf{m}}$ is rather convoluted. In fact $J[\ell^\infty]$ is the Pontrjagin dual of Tate_ℓ^*, that is,

$$J[\ell^\infty]^\wedge = \mathrm{Tate}_\ell^* = \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathrm{Tate}_\ell, \mathbf{Z}_\ell)$$

and

$$(\mathrm{Tate}_\ell^*)^\wedge = \mathrm{Hom}(\mathrm{Tate}_\ell^*, \mathbf{Q}_\ell/\mathbf{Z}_\ell) = J[\ell^\infty].$$

[[First: Why are they dual? Second: Why are we homing into $\mathbf{Q}_\ell/\mathbf{Z}_\ell$ instead of $\mathbf{Q}/\mathbf{Z}$?]] Looking at the $\mathbf{m}$-adic part shows that

$$J[\mathbf{m}^\infty] = \mathrm{Hom}(\mathrm{Tate}_{\mathbf{m}}^*, \mathbf{Q}_\ell/\mathbf{Z}_\ell)$$

and hence

$$J[\mathbf{m}] = \mathrm{Hom}(\mathrm{Tate}_{\mathbf{m}}^* / \mathbf{m}\, \mathrm{Tate}_{\mathbf{m}}^*, \mathbf{Z}/\ell\mathbf{Z}).$$

Thus if $J[\mathbf{m}] = 0$ then Nakayama's lemma implies $\mathrm{Tate}_{\mathbf{m}}^* = 0$. By autoduality this implies $\mathrm{Tate}_{\mathbf{m}} = 0$. $\square$

We have two goals. The first is to show that $t = 1$, i.e., that $J[\mathbf{m}]$ is 2-dimensional over $\mathbf{T}/\mathbf{m}$. The second is to prove that $\mathbf{T}_{\mathbf{m}}$ is Gorenstein, i.e., that $\mathbf{T}_{\mathbf{m}} \cong \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{T}_{\mathbf{m}}, \mathbf{Z}_\ell)$. This is one of the main theorems in the subject. We are assuming throughout that $\rho_{\mathbf{m}}$ is irreducible and $\ell \nmid 2N$. Loosely speaking the condition that $\ell \nmid 2N$ means that $J[\mathbf{m}]$ has good reduction at ℓ and that $J[\mathbf{m}]$ can be understood just by understanding $J[\mathbf{m}]$ in characteristic ℓ. We want to prove that $\mathbf{T}_{\mathbf{m}}$ is Gorenstein because this property plays an essential role in proving that $\mathbf{T}_{\mathbf{m}}$ is a local complete intersection.

Example 13.2.4. K. Buzzard presented the following example of a non-Gorenstein ring. Let

$$T = \{(a, b, c, d) \in \mathbf{Z}_p^4 : a \equiv b \equiv c \equiv d \pmod{p}\}.$$

Then T is local but not Gorenstein.

For now we temporarily postpone the proof of the first goal and instead show that the first goal implies the second.

Theorem 13.2.5. *Suppose $J[\mathbf{m}]$ is two dimensional over $\mathbf{T}/\mathbf{m}$ (thus $t=1$). Then $\mathbf{T}_\mathbf{m}$ is Gorenstein.*

Proof. We have seen before that

$$\begin{aligned} J[\mathbf{m}] &= \mathrm{Hom}_{\mathbf{Z}/\ell\mathbf{Z}}(\mathrm{Tate}^*_\mathbf{m}/\mathbf{m}\,\mathrm{Tate}^*_\mathbf{m}, \mathbf{Z}/\ell\mathbf{Z}) \\ &= \mathrm{Hom}_{\mathbf{T}/\mathbf{m}}(\mathrm{Tate}^*_\mathbf{m}/\mathbf{m}\,\mathrm{Tate}^*_\mathbf{m}, \mathbf{T}/\mathbf{m}). \end{aligned}$$

Thus the dual of $\mathrm{Tate}^*_\mathbf{m}/\mathbf{m}\,\mathrm{Tate}^*_\mathbf{m}$ is two dimensional over $\mathbf{T}/\mathbf{m}$ and hence $\mathrm{Tate}^*_\mathbf{m}/\mathbf{m}\,\mathrm{Tate}^*_\mathbf{m}$ itself is two dimensional over $\mathbf{T}/\mathbf{m}$. By Nakayama's lemma and autoduality of $\mathrm{Tate}_\mathbf{m}$ this implies $\mathrm{Tate}_\mathbf{m}$ is generated by 2 elements over $\mathbf{T}_\mathbf{m}$. There is a surjection

$$\mathbf{T}_\mathbf{m} \times \mathbf{T}_\mathbf{m} \twoheadrightarrow \mathrm{Tate}_\mathbf{m}.$$

In fact it is true that $\mathrm{rank}_{\mathbf{Z}_\ell}\mathrm{Tate}_\mathbf{m} = 2\,\mathrm{rank}_{\mathbf{Z}_\ell}\mathbf{T}_\mathbf{m}$. We temporarily postpone the proof of this claim. Assuming this claim and using the fact that a surjection between $\mathbf{Z}_\ell$-modules of the same rank is an isomorphism implies that $\mathrm{Tate}_\mathbf{m} \cong \mathbf{T}_\mathbf{m}\times\mathbf{T}_\mathbf{m}$. Now $\mathbf{T}_\mathbf{m}$ is a direct summand of the free $\mathbf{Z}_\ell$ module $\mathrm{Tate}_\mathbf{m}$ so $\mathbf{T}_\mathbf{m}$ is projective. A projective module over a local ring is free. Thus $\mathbf{T}_\mathbf{m}$ is free of rank 1 and hence autodual (Gorenstein). [[This argument is an alternative to Mazur's – it seems too easy... maybe I am missing something.]]

We return to the claim that

$$\mathrm{rank}_{\mathbf{Z}_\ell}\mathrm{Tate}_\mathbf{m} = 2\,\mathrm{rank}_{\mathbf{Z}_\ell}\mathbf{T}_\mathbf{m}.$$

This is equivalent to the assertion that

$$\dim_{\mathbf{Q}_\ell}\mathrm{Tate}_\mathbf{m}\otimes_{\mathbf{Z}_\ell}\mathbf{Q}_\ell = 2\dim_{\mathbf{Q}_\ell}\mathbf{T}_\mathbf{m}\otimes_{\mathbf{Z}_\ell}\mathbf{Q}_\ell.$$

The module $\mathrm{Tate}_\ell J_0(N)$ is the projective limit of the ℓ-power torsion on the Jacobian

$$J(\mathbf{C}) = \frac{\mathrm{Hom}_\mathbf{C}(S_2(\Gamma_0(N),\mathbf{C}),\mathbf{C})}{H_1(X_0(N),\mathbf{Z})}.$$

Let $L = H_1(X_0(N),\mathbf{Z})$ be the lattice. Then L is a $\mathbf{T}$-module and $\mathrm{Tate}_\ell = L\otimes_\mathbf{Z}\mathbf{Z}_\ell$ (since $L/\ell^i L \cong (\frac{1}{\ell^i}L)/L$). Tensoring with $\mathbf{R}$ gives

$$\begin{aligned} L\otimes_\mathbf{Z}\mathbf{R} &= \mathrm{Hom}_\mathbf{C}(S_2(\Gamma_0(N),\mathbf{C}),\mathbf{C}) \\ &= \mathrm{Hom}_\mathbf{R}(S_2(\Gamma_0(N),\mathbf{R}),\mathbf{C}) \\ &= \mathrm{Hom}_\mathbf{R}(S_2(\Gamma_0(N),\mathbf{R}),\mathbf{R})\otimes_\mathbf{R}\mathbf{C} = (\mathbf{T}\otimes_\mathbf{Z}\mathbf{R})\otimes_\mathbf{R}\mathbf{C} \end{aligned}$$

Thus $L\otimes_\mathbf{Z}\mathbf{R}$ is free of rank 2 over $\mathbf{T}\otimes_\mathbf{Z}\mathbf{R}$ and $L\otimes_\mathbf{Z}\mathbf{C}$ is free of rank 2 over $\mathbf{T}\otimes_\mathbf{Z}\mathbf{C}$. Next choose an embedding $\mathbf{Q}_\ell\hookrightarrow\mathbf{C}$. Now Tate_ℓ is a module over $\mathbf{T}\otimes\mathbf{Z}_\ell = \prod_{\mathbf{m}|\ell}\mathbf{T}_\mathbf{m}$ so we have a decomposition $\mathrm{Tate}_\ell = \prod_{\mathbf{m}|\ell}\mathrm{Tate}_\mathbf{m}$. Since

$$\mathrm{Tate}_\ell\otimes_{\mathbf{Z}_\ell}\mathbf{Q}_\ell = \prod_\mathbf{m}\mathrm{Tate}_\mathbf{m}\otimes_{\mathbf{Z}_\ell}\mathbf{Q}_\ell$$

we can tensor with $\mathbf{C}$ to see that

$$\mathrm{Tate}_\ell\otimes_{\mathbf{Z}_\ell}\mathbf{C} = \prod_\mathbf{m}\mathrm{Tate}_\mathbf{m}\otimes_{\mathbf{Z}_\ell}\mathbf{C}.$$

But $\mathrm{Tate}_\ell\otimes_{\mathbf{Z}_\ell}\mathbf{C} = L\otimes_\mathbf{Z}\mathbf{C}$ is free of rank 2 over $\mathbf{T}\otimes\mathbf{C}$. Therefore the product $\prod_\mathbf{m}\mathrm{Tate}_\mathbf{m}\otimes_{\mathbf{Z}_\ell}\mathbf{C}$ is free of rank 2 over $\mathbf{T}\otimes\mathbf{C}$. Since

$$\mathbf{T}\otimes\mathbf{C} = (\mathbf{T}\otimes_\mathbf{Z}\mathbf{Z}_\ell)\otimes_{\mathbf{Z}_\ell}\mathbf{C} = \prod_\mathbf{m}(\mathbf{T}_\mathbf{m}\otimes_{\mathbf{Z}_\ell}\mathbf{C})$$

we conclude that for each $\mathbf{m}$, $\mathrm{Tate}_\mathbf{m}\otimes_{\mathbf{Z}_\ell}\mathbf{C}$ is free of rank 2 over $\mathbf{T}_\mathbf{m}\otimes_{\mathbf{Z}_\ell}\mathbf{C}$. This implies

$$\dim_\mathbf{C}\mathrm{Tate}_\mathbf{m}\otimes_{\mathbf{Z}_\ell}\mathbf{C} = 2\dim_\mathbf{C}\mathbf{T}_\mathbf{m}\otimes_{\mathbf{Z}_\ell}\mathbf{C}$$

which completes the proof. □

13.2.1 Vague Comments

Ogus commented that this same proof shows that $\mathbf{T} \otimes_{\mathbf{Z}} \mathbf{C}$ is Gorenstein. Then he said that something called "faithfully flat descent" could then show that $\mathbf{T} \otimes_{\mathbf{Z}} \mathbf{Q}$ is Gorenstein.

We have given the classical argument of Mazur that $\mathbf{T}_{\mathbf{m}}$ is Gorenstein, but we still haven't shown that $J[\mathbf{m}]$ has dimension 2. This will be accomplished next time using Dieudonné modules.

13.3 Finite Flat Group Schemes

Definition 13.3.1. Let S be a scheme. Then a **group scheme over** S is a group object in the category of S-schemes.

Thus a group scheme over S is a scheme G/S equipped with S-morphisms $m : G \times G \to G$, $\mathrm{inv} : G \to G$ and a section $1_G : S \to G$ satisfying the usual group axioms.

Suppose G is a group scheme over the finite field $\mathbf{F}_q$. If R is an $\mathbf{F}_q$-algebra then $G(R) = \mathrm{Mor}(\mathrm{spec}\, R, G)$ is a group. It is the group of R-valued points of G.

We consider several standard examples of group schemes.

Example 13.3.2. The multiplicative group scheme G_m is $G_m = \mathrm{spec}\, \mathbf{Z}[x, \frac{1}{x}]$ with morphisms. [[give maps, etc.]] The additive group scheme is $\mathrm{spec}\, \mathbf{Z}[x]$...

Example 13.3.3. The group scheme $\boldsymbol{\mu}_p$ is the kernel of the morphism $G_m \to G_m$ induced by $x \mapsto x^p$. Thus $\boldsymbol{\mu}_p = \mathrm{spec}\, \mathbf{Z}[x]/(x^p - 1)$ and so for any $\mathbf{F}_q$-algebra R we have that $\boldsymbol{\mu}_p(R) = \{r \in R : r^p = 1\}$. The group scheme α_p is the kernel of the morphism $G_a \to G_a$ induced by [[what!! what is alphap?? it should be the additive group scheme of order p, no?]].

Let A be a finite algebra over $\mathbf{F}_p$ and suppose $G = \mathrm{spec}\, A$ affine group scheme (over $\mathbf{F}_p$). Then the **order** of G is defined to be the dimension of A as an $\mathbf{F}_p$ vector space.

Example 13.3.4. Let $E/\mathbf{F}_p$ be an elliptic curve. Then $G = E[p]$ is a group scheme of order p^2 [[why is this true?]]. This is wonderful because this is the order that $E[p]$ should have in analogy with the characteristic 0 situation. When we just look at points we have

$$\#G(\overline{\mathbf{F}_p}) = \begin{cases} 1 & \text{supersingular} \\ p & \text{ordinary} \end{cases}.$$

13.4 Reformulation of $V = W$ problem

Let $J = J_0(N)$ be the Jacobian of $X_0(N)$. Then J is defined over $\mathbf{Q}$ and has good reduction at all primes not dividing N. Assume ℓ is a prime not dividing N. $J[\ell]$ extends to a finite flat group scheme over $\mathbf{Z}[\frac{1}{N}]$. This is a nontrivial result of Grothendieck (SGA 7I, LNM 288). Since $\ell \nmid N$, $J[\ell]$ gives rise to a group scheme over $\mathbf{F}_\ell$.

We have "forcefully" constructed a Galois representation $\rho_{\mathbf{m}} : G \to V$ of dimension 2 over $\mathbf{T}/\mathbf{m}$. Our goal is to show that this is isomorphic to the naturally defined Galois representation $W = J[\mathbf{m}]$. So far we know that

$$0 \subset V \subset W \subset J[\ell].$$

Our assumptions are that $\ell \nmid N$, V is irreducible, and $\ell \neq 2$.

Let $\underline{J}$ be J thought of as a scheme over $\mathbf{Z}_\ell$. Grothendieck showed that $\underline{J}$ is the spectrum of a free finite $\mathbf{Z}_\ell$-module. Raynaud (1974) showed that if $\ell \neq 2$ then essentially everything about $\underline{J}[\ell]$ can be seen in terms of $J[\ell](\overline{Q}_\ell)$. He goes on to construct group scheme $\underline{V}$ and $\underline{W}$ over $\mathbf{Z}_\ell$ such that

$$\underline{V} \subset \underline{W} \subset \underline{J}[\ell].$$

Our goal is to prove that the inclusion $V \hookrightarrow W$ of Galois modules is an isomorphism. Raynaud showed noted that the category of finite flat group schemes over $\mathbf{Z}_\ell$ is an abelian category so the cokernel $\underline{Q} = \underline{W}/\underline{V}$ is defined. Furthermore, $V = W$ iff $\underline{Q} = 0$. Since $\underline{Q}$ is flat $\underline{Q}_{\mathbf{F}_\ell}$ has the same dimension over $\mathbf{F}_\ell$ as $\underline{Q}_{\mathbf{Q}_\ell}$ has over $\mathbf{Q}_\ell$. Passing to characteristic ℓ yields an exact sequence

$$0 \to \underline{V}_{\mathbf{F}_\ell} \to \underline{W}_{\mathbf{F}_\ell} \to \underline{Q}_{\mathbf{F}_\ell} \to 0.$$

Thus $V \hookrightarrow W$ is an isomorphism iff $\underline{V}_{\mathbf{F}_\ell} \hookrightarrow \underline{W}_{\mathbf{F}_\ell}$ is an isomorphism. Since $\underline{V}$, $\underline{W}$, and $\underline{Q}$ have an action of $k = \mathbf{T}/\mathbf{m}$ that are k-vector space schemes. This leads us to Dieudonné theory.

13.5 Dieudonné Theory

Let G/k be a finite k-vector space scheme where k is a finite field of order q. Suppose G has order q^n so G is locally the spectrum of a rank n algebra over k. The Dieudonné functor contravariantly associates to G an n dimensional k-vector space $D(G)$. Let $\mathrm{Frob} : G \to G$ be the morphism induced by the pth power map on the underlying rings and let Ver be the dual of Frob. Let $\varphi = D(\mathrm{Frob})$ and $\nu = D(\mathrm{Ver})$, then it is a property of the functor D that $\varphi \circ \nu = \nu \circ \varphi = 0$. The functor D is a fully faithful functor.

Example 13.5.1. Let $k = \mathbf{F}_p$. If G is either $\boldsymbol{\mu}_p$, α_p, or $\mathbf{Z}/p\mathbf{Z}$ then $D(G)$ is a one-dimensional vector space over $\mathbf{F}$. In the case of α_p, $\varphi = \nu = 0$. For $\boldsymbol{\mu}_p$, $\varphi = 0$ and $\nu \neq 1$ and for $\mathbf{Z}/p\mathbf{Z}$, $\varphi \neq 1$ and $\nu = 0$. [[The latter two could be reversed!]]

Let $G^\vee = \mathrm{Hom}(G, \boldsymbol{\mu}_p)$ denote the Cartier dual of the scheme G. Then

$$D(G^\vee) = \mathrm{Hom}_k(D(G), k)$$

(φ and ν are switched.)

Example 13.5.2. Let A be an abelian variety over $\mathbf{F}_\ell$ and let $G = A[\ell]$. Then G is an $\mathbf{F}_\ell$-vector space scheme of order ℓ^{2g}. Thus $D(G)$ is a $2g$-dimensional $\mathbf{F}_\ell$-vector space and furthermore $D(G) = H^1_{DR}(A/\mathbf{F}_\ell)$. The Hodge filtration on H^1_{DR} of the abelian variety A gives rise to a diagram

$$\begin{array}{ccccccccccc}
 & & & & & & \mathrm{Hom}(H^0(A^\vee,\Omega^1),\mathbf{F}_\ell) & = & Tan(A^\vee) & & \\
 & & & & & & & & \| & & \\
0 & \to & H^0(A,\Omega^1) & \to & D(G) & \to & H^1_{DR}(A/\mathbf{F}_\ell) & \to & H^1(A,\mathcal{O}) & \to & 0 \\
 & & & & & & \| & & \| & & \\
 & & & & & & D(A_{/\mathbf{F}_\ell}[\ell]) & \to & D(A_{/\mathbf{F}_\ell}[\mathrm{Ver}]) & \to & 0
\end{array}$$

There is an exact sequence

$$0 \to W_{\mathbf{F}_\ell} \to J_{\mathbf{F}_\ell}[\ell] \xrightarrow{\mathbf{m}} J_{\mathbf{F}_\ell}[\ell]$$

so because D is an exact functor the sequence

$$D(J_{\mathbf{F}_\ell}[\ell]) \xrightarrow{\mathbf{m}} D(J_{\mathbf{F}_\ell}[\ell]) \to D(W_{\mathbf{F}_\ell}) \to 0$$

is exact. Following Fontaine we consider

$$D(W_{\mathbf{F}_\ell}[\mathrm{Ver}]) = H^1(J,\mathcal{O})/\mathbf{m}H^1(J,\mathcal{O}).$$

13.6 The Proof: Part II

[[We all just returned from the Washington D.C. conference and will now resume the proof.]]

Let $J_0(N)$ be the Jacobian of $X_0(N)$. Let $\mathbf{m} \subset \mathbf{T}$ be a maximal ideal and suppose $\mathbf{m}|\ell$. Assume that $\ell \neq 2$ and $\ell \nmid N$. The assumption that $\ell \neq 2$ is necessary for Raynaud's theory and we assume that $\ell \nmid N$ so that our group schemes will have good reduction. We attach to $\mathbf{m}$ a 2 dimensional semisimple representation $\rho_{\mathbf{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to V$ and only consider the case that $\rho_{\mathbf{m}}$ is irreducible.

The $\mathbf{m}$-torsion of the Jacobian, $W = J_0(\overline{\mathbf{Q}})[\mathbf{m}]$, is naturally a Galois module. We have shown that $W \neq 0$. By (Boston-Lenstra-Ribet) $W \cong V \times \cdots \times V$ (the number of fractions is not determined). We proved that $W^{s.s.} \cong V \times \cdots \times V$. Choose an inclusion $V \hookrightarrow W$ and let $Q = W/V$ be the cokernel.

Theorem 13.6.1. *$Q = 0$ so* $\dim_{\mathbf{T}/\mathbf{m}} W = 2$

To prove the theorem we introduce the "machine" of finite flat group schemes over $\mathbf{Z}_\ell$. For example, W extends to a finite flat group scheme $W_{\mathbf{Z}_\ell}$ which is defined to be the Zariski closure of W in $J_{\mathbf{Z}_\ell}[\ell]$. Passing to group schemes yields an exact sequence

$$0 \to V_{\mathbf{Z}_\ell} \to W_{\mathbf{Z}_\ell} \to Q_{\mathbf{Z}_\ell} \to 0.$$

Reducing mod ℓ then yields an exact sequence of $\mathbf{F}_\ell$-group schemes

$$0 \to V_{\mathbf{F}_\ell} \to W_{\mathbf{F}_\ell} \to Q_{\mathbf{F}_\ell} \to 0.$$

The point is that $Q = 0$ iff $Q_{\mathbf{Z}_\ell} = 0$ iff $Q_{\mathbf{F}_\ell} = 0$.

Next we introduced the exact contravariant Dieudonné functor

$$D : (\text{ Groups Schemes } /\mathbf{F}_\ell \text{ }) \longrightarrow (\text{ Linear Algebra }).$$

D sends a group scheme G to a $\mathbf{T}/\mathbf{m}$ vector space equipped with two endomorphisms $\varphi = \mathrm{Frob}$ and $\nu = \mathrm{Ver}$. Applying D gives an exact sequence of $\mathbf{T}/\mathbf{m}$-vector spaces

$$0 \to D(Q) \to D(W) \to D(V) \to 0$$

where everything is now viewed over $\mathbf{F}_\ell$.

Lemma 13.6.2. $D(W[\mathrm{Ver}]) = (H^0(X_0(N)_{\mathbf{F}_\ell}, \Omega^1)[\mathbf{m}])^*$

Proof. We have the diagram

$$\begin{array}{ccc} D(J_{\mathbf{F}_\ell}[\ell]) & = & H^1_{DR}(J_{\mathbf{F}_\ell}) \\ \downarrow & & \downarrow \\ D(J_{\mathbf{F}_\ell}[\mathrm{Ver}]) & = & H^1(J_{\mathbf{F}_\ell}, \mathcal{O}_{J_{\mathbf{F}_\ell}}) \\ \downarrow & & \downarrow \\ D(W[\mathrm{Ver}]) & = & H^1(J_{\mathbf{F}_\ell}, \mathcal{O}_J)/\mathbf{m}H^1(J_{\mathbf{F}_\ell}, \mathcal{O}_J) \end{array}$$

Furthermore we have the identifications

$$\begin{aligned} H^1(J_{\mathbf{F}_\ell}, \mathcal{O}_J) &= \mathrm{Tan}(J^\vee_{\mathbf{F}_\ell}) = \mathrm{Cot}(J^\vee_{\mathbf{F}_\ell})^* \\ &= H^0(J^\vee, \Omega^1)^* = H^0(X_0(N), \Omega^1)^* \end{aligned}$$

For the last identification we must have $J^\vee = \mathrm{Alb}(X_0(N))$. Finally

$$\begin{aligned} D(W[\mathrm{Ver}]) &= H^1(J, \mathcal{O}_J)/\mathbf{m}H^1(J, \mathcal{O}_J) \\ &= (H^0(X_0(N)_{\mathbf{F}_\ell}, \Omega^1)[\mathbf{m}])^*. \end{aligned}$$

□

Lemma 13.6.3. *$H^0(X_0(N)_{\mathbf{F}_\ell}, \Omega^1)[\mathbf{m}]$ has $\mathbf{T}/\mathbf{m}$ dimension ≤ 1.*

Proof. Let $S = H^0(X_0(N)_{\mathbf{F}_\ell}, \Omega^1)[\mathbf{m}]$. Then $S \hookrightarrow \mathbf{F}_\ell[[q]]$. We defined the Hecke operators T_n on S via the identification $S \cong H^1(J, \mathcal{O}_J)$ so that they act on $S \subset \mathbf{F}_\ell[[q]]$ in the standard way. Let $\mathbf{T}(S)$ be the subalgebra of $\mathrm{End}(S)$ generated by the images of the T_n in $\mathrm{End}(S)$. ($\mathbf{T}(S)$ is not a subring of $\mathbf{T}$.) There is a perfect pairing

$$\begin{aligned}\mathbf{T}(S) \times S &\longrightarrow \mathbf{F}_\ell\\ (T, f) &\mapsto a_1(f|T)\end{aligned}$$

Thus $\dim_{\mathbf{F}_\ell} \mathbf{T}(S) = \dim_{\mathbf{F}_\ell} S$ and so $\dim_{\mathbf{T}(S)} S \leq 1$. Since $\mathbf{m}$ acts trivially on S there is a surjection $\mathbf{T}/\mathbf{m} \twoheadrightarrow \mathbf{T}(S)$. Thus

$$\dim_{\mathbf{T}/\mathbf{m}} S \leq \dim_{\mathbf{T}(S)} S \leq 1$$

as desired. ∎

An application of the above lemma shows that $D(W[\mathrm{Ver}])$ has $\mathbf{T}/\mathbf{m}$ dimension ≤ 1.

Lemma 13.6.4. *$D(W[\mathrm{Ver}]) \cong D(V[\mathrm{Ver}])$*

Proof. Consider the following diagram.

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & D(Q) & \to & D(W) & \to & D(V) & \to & 0\\
 & & \downarrow \mathrm{Ver} & & \downarrow \mathrm{Ver} & & \downarrow \mathrm{Ver} & & \\
0 & \to & D(Q) & \to & D(W) & \to & D(V) & \to & 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & D(Q)/\mathrm{Ver}\, D(Q) & \to & D(W)/\mathrm{Ver}\, D(W) & \xrightarrow{?\sim?} & D(V)/\mathrm{Ver}\, D(V) & \to & 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

$D(V)$ has dimension 2 so since $\mathrm{Ver} \circ \mathrm{Frob} = \mathrm{Frob} \circ \mathrm{Ver} = 0$ and Ver, Frob are both nonzero they must each have rank 1 (in the sense of undergraduate linear algebra). Since D is exact, $D(V[\mathrm{Ver}]) = D(V)/\mathrm{Ver}\, D(V)$ and $D(W[\mathrm{Ver}]) = D(W)/\mathrm{Ver}\, D(W)$. By the previous lemma $\dim D(W)/\mathrm{Ver}\, D(W) = 1$. Thus $D(W)/\mathrm{Ver}\, D(W) \to D(V)/\mathrm{Ver}\, D(V)$ is a map of 1 dimensional vector spaces so to show that it is an isomorphism we just need to show that it is surjective. This follows from the commutativity of the square

$$\begin{array}{ccccc}
D(W) & \to & D(V) & \to & 0\\
\downarrow & & \downarrow & & \\
D(W)/\mathrm{Ver}\, D(W) & \to & D(V)/\mathrm{Ver}\, D(V) & & \\
 & & \downarrow & & \\
 & & 0 & &
\end{array}$$

∎

Suppose for the moment that we admit (Boston-Lenstra-Ribet). Then

$$W = V \times \cdots \times V = V^{\oplus t}$$

so

$$D(W[\mathrm{Ver}]) \cong D(V[\mathrm{Ver}])^{\oplus t}$$

and hence $t = 1$.

Alternatively we can avoid the use of Boston-Lenstra-Ribet. Suppose $Q \neq 0$. Then there is an injection $V \hookrightarrow Q$. [[I can't see this without using B-L-R. It isn't obvious to me from $0 \to V \to W \to Q \to 0$.]] Thus over $\mathbf{F}_\ell$, $V[\mathrm{Ver}] \hookrightarrow Q[\mathrm{Ver}]$. Since $V[\mathrm{Ver}] \neq 0$ this implies $Q[\mathrm{Ver}] \neq 0$. Thus $D(Q)/\operatorname{Ver} D(Q) = D(Q[\mathrm{Ver}]) \neq 0$. But the bottom row of the above diagram implies $D(Q)/\operatorname{Ver} D(Q) = 0$ so $Q = 0$.

13.7 Key Result of Boston-Lenstra-Ribet

Let G be a group (i.e., $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$), let k be a field (i.e., $k = \mathbf{T}/\mathbf{m}$), and let V be a two dimensional k-representation of G given by

$$\rho : k[G] \to \mathrm{End}_k(V) = M_2(k).$$

The key hypothesis is that V is absolutely irreducible, i.e., that ρ is surjective. For each $g \in G$ consider

$$p_g = g^2 - g \operatorname{Tr} \rho(g) + \det \rho(g) \in k[G].$$

By the Cayley-Hamilton theorem $\rho(p_g) = 0$. Let J be the two-sided ideal of $k[G]$ generated by all p_g such that $g \in G$. Since $J \subset \ker \rho$, ρ induces a map

$$\sigma : k[G]/J \to \mathrm{End}_k(V).$$

Theorem 13.7.1 (Boston-Lenstra-Ribet). *If σ is surjective then σ is an isomorphism.*

In particular if V is absolutely irreducible then σ is surjective. The theorem can be false when $\dim V > 2$.

Suppose W is a second representation of G given by $\mu : k[G] \to \mathrm{End}(W)$ and that $\mu(J) = \{0\} \subset \mathrm{End}(W)$. Then W is a module over $k[G]/J = \mathrm{End}(V)$. But $\mathrm{End}(V)$ is a semisimple ring so any $\mathrm{End}(V)$ module is a direct sum of simple $\mathrm{End}(V)$ modules. The only simple $\mathrm{End}(V)$ module is V. Thus $W \cong V^{\oplus n}$ for some n.

Chapter 14

Local Properties of ρ_λ

Let f be a newform of weight 2 on $\Gamma_0(N)$. To f we have associated an abelian variety $A = A_f$ furnished with an action of $E = \mathbf{Q}(\ldots, a_n(f), \ldots)$. Let $\mathcal{O}_E$ be the ring of integers of E and $\lambda \subset \mathcal{O}_E$ a prime. Then we obtain a λ-adic representation

$$\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(E_\lambda)$$

on the Tate module $\mathrm{Tate}_\ell A = \varprojlim A[\lambda^i]$. We will study the local properties of ρ_λ at various primes p.

14.1 Definitions

To view ρ_λ locally at p we restrict to the decomposition group $D_p = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ at p. Recall the definition of D_p. Let K be a finite extension of $\mathbf{Q}$ and let w be a prime of K lying over p. Then the decomposition group at w is defined to be

$$D_w = \{\sigma \in \mathrm{Gal}(K/\mathbf{Q}) : \sigma w = w\}.$$

Proposition 14.1.1. $D_w \cong \mathrm{Gal}(K_w/\mathbf{Q}_p)$

Proof. Define a map $\mathrm{Gal}(K_w/\mathbf{Q}_p) \to D_w$ by $\sigma \mapsto \sigma_{|K}$. Since $\sigma_{|K}$ fixes $\mathbf{Q}$ this restriction is an element of $\mathrm{Gal}(K/Q)$. Since $w\mathcal{O}_{K_w}$ is the unique maximal ideal of $\mathcal{O}_{K_w}$ and σ induces an automorphism of $\mathcal{O}_{K_w}$, it follows that $\sigma(w\mathcal{O}_{K_w}) = w\mathcal{O}_{K_w}$. Thus $\sigma_{|K}(w) = w$ so $\sigma_{|K} \in D_w$. The map $\sigma \mapsto \sigma_{|K}$ is bijective because K is dense in K_w. $\square$

Let

$$D_p = \varprojlim{}_{w|p} D_w = \varprojlim{}_{w|p} \mathrm{Gal}(K_w/\mathbf{Q}_p) = \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p).$$

For each $w|p$ let the inertia group I_w be the kernel of the map from D_w into $\mathrm{Gal}(\mathcal{O}_K/w, \mathbf{Z}/p\mathbf{Z})$. Let $I_p = \varprojlim{}_{w|p} I_w$.

14.2 Local Properties when $p \nmid N$

Next we study local properties of ρ_λ at primes $p \nmid N$. Thus suppose $p \nmid N$ and $p \neq \ell = \mathrm{char}(\mathcal{O}_E/\lambda)$. Let $D_p = \mathrm{Gal}(\overline{\mathbf{Q}_p}/\mathbf{Q}_p)$. Then

1) $\rho_\lambda|D_p$ is unramified (i.e., $\rho_\lambda(I_p) = \{1\}$) thus $\rho_\lambda|D_p$ factors through D_p/I_p so $\rho_\lambda(\mathrm{Frob}_p)$ is defined.

2) $\mathrm{Tr}(\rho_\lambda(\mathrm{Frob}_p)) = a_p(f)$

2+) We can describe $\rho_\lambda|D_p$ up to isomorphism. It is the unique semisimple representation satisfying 1) and 2).

14.3 Weil-Deligne Groups

Notice that everything is sort of independent of λ. Using Weil-Deligne groups we can summarize all of these λ-adic representations in terms of data which makes λ disappear.

We have an exact sequence

$$1 \to I_p \to \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p) \to 0.$$

Since $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p) = \hat{\mathbf{Z}}$ there is an injection $\mathbf{Z} \hookrightarrow \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$. Define the Weil group $W(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \subset \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ to be the set of elements of $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ mapping to $\mathbf{Z} \subset \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$. W fits into the exact sequence

$$1 \to I_p \to W(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathbf{Z} \to 1.$$

There is a standard way in which the newform f gives rise to a representation of W. Factor the polynomial $x^2 - a_p(f)x + p$ as a product $(x-r)(x-r')$ with $r, r' \in \mathbf{C}$. Define maps α, β by

$$\alpha : \mathbf{Z} \to \mathbf{C}^* : 1 \mapsto r$$

$$\beta : \mathbf{Z} \to \mathbf{C}^* : 1 \mapsto r'$$

Combining α and β and the map $W(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathbf{Z}$ yields a map

$$\alpha \oplus \beta : W(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{GL}_2(\mathbf{C})$$

$$\sigma \mapsto \begin{pmatrix} \alpha(\sigma) & 0 \\ 0 & \beta(\sigma) \end{pmatrix}.$$

Moreover $\alpha \oplus \beta$ gives rise via some construction to all the λ-adic representations ρ_λ.

14.4 Local Properties when $p|N$

Suppose $p|N$ but $p \neq \ell$. Carayol was able to generalize 1) in his thesis which builds upon the work of Langlands and Deligne in the direction of Deligne-Rapaport and Katz-Mazur. The idea is that the abelian variety A has a conductor which is a positive integer divisible by those primes of bad reduction. The conductor of A satisfies $\mathrm{cond}(A) = M^g$ where $g = \dim A$ and M is the reduced conductor of A.

Theorem 14.4.1 (Carayol). *$M = N$.*

How can we generalize 2) or 2+)? For each p dividing N there is a representation σ_p of $\mathrm{WD}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ over $\mathbf{C}$ such that σ_p gives rise to $\rho_\lambda|D_p$ for all $\lambda \nmid p$. Here $\mathrm{WD}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ is the Weil-Deligne group which is Deligne's generalization of the Weil group. $\mathrm{WD}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ is an extension of $W(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$. What is σ_p supposed to be? The point is that σ_p is determined by f. Thus every f gives rise to a family $(\sigma_p)_{p \text{ prime}}$. To really think about σ_p we must think about modular forms in an adelic context instead of viewing them as holomorphic functions on the complex upper halfplane.

If $p^2|N$ and $f = \sum a_n q^n$ is a newform of level p^2 then it is a classical fact that $a_p = 0$. But the study of $\rho_\lambda|D_p$ is rich and "corresponds to a rather innocuous looking crystal".

14.5 Definition of the Reduced Conductor

We now define the reduced conductor. Let λ be a prime of E and p a prime of $\mathbf{Q}$ such that $\lambda \not| p$. We want to define some integer $e(p)$ so that $p^{e(p)}$ is the p-part of the reduced conductor. We will not define $e(p)$ but what we will do is define an integer $e(p,\lambda)$ which is the p part of the conductor. $e(p,\lambda)$ is independent of λ but this will not be proved here.

Consider

$$\rho_\lambda | D_p : \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{Aut}_{E_\lambda} V.$$

Let $V^I \subset V$ be the inertia invariants of V, i.e.,

$$V^I = \{v \in V : \rho_\lambda(\sigma)(v) = v \text{ for all } \sigma \in I\}.$$

Since ρ_λ is unramified at p iff $\rho_\lambda(I_p) = \{1\}$ we comment that ρ_λ is unramified at p iff $V^I = V$. Let

$$e(p,\lambda) = \dim V/V^I + \delta(p,\lambda)$$

where $\delta(p,\lambda)$ is the Swan conductor. By working with finite representations we define $\delta(p,\lambda)$ as follows. Choose a D_p-stable lattice L in V by first choosing an arbitrary one then taking the sum of its finitely many conjugates. Let $\overline{V} = L/\lambda L$ which is a 2 dimensional vector space over $k = \mathcal{O}_E/\lambda$. Let G be the quotient of $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ by the kernel of the map $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) \to \mathrm{Aut}_K \overline{V}$. Thus $G = \mathrm{Gal}(K/\mathbf{Q}_p)$ for some finite extension $K/\mathbf{Q}_p$. The extension K is finite over $\mathbf{Q}_p$ since $G \subset \mathrm{Aut}_k \overline{V}$ and $\mathrm{Aut}_k \overline{V}$ is a 2×2 matrix ring over a finite field. The corresponding diagram is

$$\begin{array}{ccc} \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p) & \to & \mathrm{Aut}_k \overline{V} \\ \searrow & & \cup \\ & & G = \mathrm{Gal}(K/\mathbf{Q}_p) \end{array}$$

Consider in G the sequence of "higher ramification groups"

$$G = G_{-1} \supset G_0 \supset G_1 \supset \cdots .$$

Here G_0 is the inertia group of $K/\mathbf{Q}_p$ and G_1 is the p-sylow subgroup of G_0 [[the usages of "the" in this sentence makes me nervous.]] Let $G_i = \{g \in G_0 : \mathrm{ord}(g\pi - \pi) \geq i+1\}$ where π is some kind of uniformizing parameter [[I missed this – what is π?]] Let

$$\delta(p,\lambda) = \sum_{i=1}^{\infty} \frac{1}{(G_0 : G_i)} \dim(\overline{V}/\overline{V}^{G_i}).$$

It is a theorem that $\delta(p,\lambda)$ is an integer and does not depend on λ.

If to start with we only had $\overline{V}$ and not V we could define

$$\mathrm{cond}(\overline{V}) = \sum_{i=0}^{\infty} \frac{1}{(G_0 : G_i)} \dim(\overline{V}/\overline{V}^{G_i}).$$

Then

$$e(p,\lambda) = \mathrm{cond}(\overline{V}) + (\dim \overline{V}^I - \dim V^I).$$

A reference for much of this material is Serre's *Local Factors of L-functions of λ-adic Representations.*

14.6 Introduction

Our goal is to study local properties of the λ-adic representations ρ_λ arising from a weight 2 newform of level N on $\Gamma_0(N)$. There is a theorem of Carayol which states roughly that if $p \neq \ell$ then $\rho_\lambda | D_p$ is predictable from "the component at p of f". To understand this theorem we must understand what is meant by "the component at p of f". If $p^2 \nmid N$ this component is easy to determine but if $p^2 | N$ it is harder. One reason is that when $p^2|N$ then $a_p(f) = 0$. [[this should be easy to see so there should be an argument here.]] If $a_p(f) = 0$ and $\rho_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(V_\lambda)$ then $V_\lambda^{I_p} = 0$. This means that there is no ramification going on at p. See Casselman, "On representations of $GL(2)$ and the arithmetic of modular curves", Antwerp II.

14.7 Adelic Representations Associated to Modular Forms

Let R be a subring of $\mathbf{A}^2 = \mathbf{R}^2 \times (\hat{\mathbf{Z}} \otimes_{\mathbf{Z}} \mathbf{Q})^2$, suppose that $R \cong \mathbf{Q}^2$ and that $\mathbf{R} \otimes_{\mathbf{Q}} \mathbf{A} \cong \mathbf{A}^2$. Let $L = R \cap (\mathbf{R}^2 \times \hat{\mathbf{Z}}^2)$. Then the natural map $L \otimes \hat{\mathbf{Z}} \to \mathbf{Z}^2$ is an isomorphism. [[Is the isomorphism implied by the definition of L or is it part of the requirement for L to actually form an adelic lattice?]] L is called an *adelic lattice*.

The space of modular forms $S_2(\Gamma_0(N))$ is isomorphic to a certain space of functions on $G(\mathbf{A}) = \mathrm{GL}(2, \mathbf{A})$. See Borel-Jacquet [[Corvallis?]] or Diamond-Taylor, Inventiones Mathematica, 115 (1994) [[what is title?]] We will describe this isomorphism.

Write $\mathbf{A} = \mathbf{R} \times \mathbf{A}_f = \mathbf{A}_\infty \times \mathbf{A}^\infty$ where $\mathbf{A}_f = \prod \mathbf{Q}_p$ (restricted product) is the ring of finite adeles. $\mathbf{A}_\infty$ denotes the adeles with respect to the place ∞ so $\mathbf{A}_\infty = \mathbf{R}$, and $\mathbf{A}^\infty$ denotes the adeles away from the place ∞ so $\mathbf{A}^\infty = \mathbf{A}_f$.

$S_2(\Gamma_0(N))$ is isomorphic to the set of functions $\varphi : G(\mathbf{A}) \to \mathbf{C}$ which satisfy

0) φ is left invariant by $G(\mathbf{Q})$, i.e., $\varphi(x) = \varphi(gx)$ for all $g \in G(\mathbf{Q})$ and all $x \in G(\mathbf{A})$,

1) $\varphi(xu^\infty) = \varphi(x)$ for all $x \in G(\mathbf{A})$ and all $u^\infty \in U^\infty$,

2) $\varphi(xu_\infty) = \varphi(x) j(u_\infty, i)^{-k} \det(u_\infty)$ for all $x \in G(\mathbf{A})$ and $u_\infty \in U_\infty$,

3) holomorphy, cuspidal, and growth conditions.

U^∞ is the adelic version of $\Gamma_0(N)$. Thus U^∞ is the compact open subgroup

$$U^\infty = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\hat{\mathbf{Z}}) : c \equiv 0 \pmod{N}\} \subset G(\mathbf{A}_f).$$

(For $\Gamma_1(N)$ the condition is that $c \equiv 0 \pmod{N}$ and $d \equiv 1 \pmod{N}$ but a is not restricted.)

Next we describe $U_\infty \subset \mathrm{GL}_2(\mathbf{R})$. $\mathrm{GL}_2(\mathbf{R})$ operates on $\mathcal{H}^\pm = \mathbf{C} - \mathbf{R}$ by $z \mapsto \frac{az+b}{cz+d}$. Let U_∞ be the stabilizer of i.

The third condition involves the *automorphy factor* j defined by

$$j(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, z) = cz + d.$$

To explain the holomorphy condition 3) we define, for any $g \in G(\mathbf{A}_f)$ a map

$$\alpha_{g,\varphi} : \mathcal{H}^\pm \to \mathbf{C}$$

$$hi \mapsto \varphi(gh) j(h,i)^k (\det(h))^{-1}.$$

Here $h \in G(\mathbf{A}_f)$ so $hi \in \mathcal{H}^\pm$. There may be several different $h \in G(\mathbf{A}_f)$ which give the same $hi \in \mathcal{H}^\pm$ so it must be checked that $\alpha_{g,\varphi}$ is well-defined. Suppose $hi = h'i$, then $h^{-1}h'i = i$. Thus $h^{-1}h' \in U_\infty$, so by 2),

$$\varphi(gh^{-1}h') = \varphi(g)j(h^{-1}h', i)^{-k}\det(h^{-1}h').$$

Thus

$$\varphi(gh^{-1}h')\det(h')^{-1} = \varphi(g)j(h^{-1}h', i)^{-k}\det(h)^{-1}.$$

Substituting gh for g yields

$$\varphi(gh')(\det(h'))^{-1} = \varphi(gh)j(h^{-1}h', i)^{-k}(\det(h))^{-1}.$$

[[Why does not the automorphy factor work out right?? :(]] The holomorphy condition is that the family of maps $\alpha_{g,\varphi}$ are all holomorphic.

The cuspidal condition is that for all $g \in G(\mathbf{A})$, the integral

$$\int_{u\in\mathbf{A}/\mathbf{Q}} \varphi(\left(\begin{smallmatrix}1 & u\\ 0 & 1\end{smallmatrix}\right), g)du = 0$$

vanishes. Since $\mathbf{A}/\mathbf{Q}$ is compact it has a Haar measure defined modulo k^* which induces du. Although the integral is not well-defined the vanishing or non-vanishing of the integral is.

We can now describe the isomorphism between $S_k(\Gamma_0(N))$ and the space of such functions φ on $G(\mathbf{A})$.

$$\{\text{ space of } \varphi \text{ satisfying 0-3}\} \to S_k(\Gamma_0(N))$$

$$\varphi \mapsto f$$

where f is the restriction to $\mathcal{H} = \mathcal{H}^+$ of the function

$$hi \mapsto \varphi(h)j(h, i)^k(\det h)^{-1}.$$

Now we can associate to a newform f a representation of $G(\mathbf{A})$. We can weaken condition 1) to get

1-) $\varphi(xu^\infty) = \varphi(x)$ for all $x \in G(\mathbf{A})$ and all $u^\infty \in U^\infty$ where U^∞ is *some* compact open subgroup of $G(\mathbf{A}_f)$.

[[can the U^∞ vary for each $x \in G(\mathbf{A})$ or are they fixed throughout?]] Let $\mathcal{S}$ be the space of all functions satisfying all conditions except 1-) replaces 1). This space has a left action of $G(\mathbf{A})$:

$$(g * \varphi)(x) = \varphi(xg)$$

If f is a newform corresponding to some φ via the above isomorphism then via this action f gives rise to an infinite dimensional representation π of $G(\mathbf{A})$. In fact we obtain, for each prime p, a representation π_p of $\mathrm{GL}(2, \mathbf{Q}_p)$. The representation space is $\sum_{g\in\mathrm{GL}_2(\mathbf{Q}_p)} \mathbf{C}\cdot g * \varphi$. Our immediate goal is to understand π_p for as many p as possible. [[spherical representations have something to do with this. are the π_p spherical reps?]]

We are studying local properties of the λ-adic representations ρ_λ associated to a newform f of weight 2, level N and character $\varepsilon : (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$ (with $\mathrm{cond}(\varepsilon)|N$). Let $\ell \in \mathbf{Z}$ be the prime over which λ lies. We look at ρ_λ locally at p, $p \neq \ell$.

As we saw last time f gives rise to an irreducible representation of $\mathrm{GL}(2, \mathbf{A})$. An irreducible representation of $GL(2, \mathbf{A})$ gives rise to a family of representations (π_v) where v is a prime or ∞ and π_v is an irreducible representation of $\mathrm{GL}(2, \mathbf{Q}_v)$. This is because $\mathbf{Q}_v \subset \mathbf{A}$ so $\mathrm{GL}(2, \mathbf{Q}_v) \subset \mathrm{GL}(2, \mathbf{A})$.

Carayol proved that if $p \neq \ell$ then $\rho_\lambda|_{D_p}$ depends, up to isomorphism, only on π_p. The most interesting case of the proof of this theorem is the case when $p^2 \not| N$. Everything necessary to obtain a proof in this case was available in the work of Langlands [1972].

To get an idea of what is going on we will first consider the case when $p \not| N$. The characteristic polynomial of Frobenious (at least psychologically) under the representation $\rho_\lambda|_{D_p}$ is

$$x^2 - a_p x + p\varepsilon(p) = (x - r)(x - s).$$

Because of Weil's proof of the Riemann hypothesis for abelian varieties [over finite fields?] one knows that $|r| = |s| = \sqrt{p}$. Since ρ_λ arises from the action of Galois on an abelian variety which has good reduction at p (since $p \not| N$) it follows that $\rho_\lambda|_{D_p}$ is unramified. [[Is this in Serre-Tate, 1968?]] We also know that $\rho_\lambda(\mathrm{Frob}_p)$ has characteristic polynomial

$$x^2 - a_p x + p\varepsilon(p) \in E[x].$$

In this situation one also knows that $\rho_\lambda(\mathrm{Frob}_p)$ is semisimple [[proof: Ribet nodded at Coleman who smiled at nodded back.]] Thus

$$\rho_\lambda \sim \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix}.$$

In this situation what is the representation π_p of $\mathrm{GL}(2, \mathbf{Q}_p)$? There are two characters

$$\alpha, \beta : \mathbf{Q}_p^* \to \mathbf{C}^*$$

(called "Grössencharacters of type (a,0)") such that

1. α and β are unramified in the sense that
$$\alpha|_{\mathbf{Z}_p^*} = \beta|_{\mathbf{Z}_p^*} = 1.$$
This is a reasonable condition since under some sort of local class field theory $\mathbf{Q}_p^*$ embeds as a dense subgroup of $\mathrm{Gal}(\overline{\mathbf{Q}_p}/\mathbf{Q}_p)$ and under this embedding the inertia subgroup $I \subset \mathrm{Gal}(\overline{\mathbf{Q}_p}/\mathbf{Q}_p)$ corresponds to $\mathbf{Z}_p^*$. [[This could be wrong. Also, is $I \cap \mathbf{Q}_p^* = \mathbf{Z}_p^*$?]]

2. $\alpha(p^{-1}) = r$, $\beta(p^{-1}) = s$.

In the 1950's Weil found a way under which α and β correspond to continuous characters α_λ, β_λ on $\mathrm{Gal}(\overline{\mathbf{Q}_p}/\mathbf{Q}_p)$ with values in $\overline{E}_\lambda^*$ such that

1. α_λ and β_λ are unramified.

2. $\alpha_\lambda(\mathrm{Frob}_p) = r$ and $\beta_\lambda(\mathrm{Frob}_p) = s$.

One has that $\rho_\lambda = \alpha_\lambda \oplus \beta_\lambda$. See Serre-Tate [29]. [[Why see this?]]

Define a character Θ on the Borel subgroup

$$B = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \subset \mathrm{GL}(2, \mathbf{Q}_p)$$

by

$$\Theta\begin{pmatrix} x & y \\ 0 & z \end{pmatrix} = \alpha(x)\beta(z) \in \mathbf{C}^*.$$

Then

$$\pi_p = \mathrm{Ind}_B^{\mathrm{GL}(2,\mathbf{Q}_p)} := \mathbf{C}[\mathrm{GL}(2,\mathbf{Q}_p)] \otimes_{\mathbf{C}[B]} \mathbf{C}.$$

We call this induced representation π_p the unramified principal series representation associated to α and β and write $\pi_p = \mathrm{PS}(\alpha,\beta)$. People say π_p is spherical in the sense that there is a vector in the representation space invariant under the maximal compact subgroup of $\mathrm{GL}(2,\mathbf{Q}_p)$. [[Ribet was slightly unsure about the correct definition of spherical.]] [[For some mysterious reason]] since $\pi_p = \mathrm{PS}(\alpha,\beta)$ it follows that α, β and hence $\rho_\lambda|D_p$ is completely determined by π_p. [[This is the point and i don't see this.]]

Next we consider the more difficult case when $p||N$ (p divides N exactly). There are two cases to consider

a ε is ramified at p ($p|\,\mathrm{cond}(\varepsilon)$)

b ε is unramified at p ($p\nmid\mathrm{cond}(\varepsilon)$).

14.8 More Local Properties of the ρ_λ.

Let f be a newform of level N. Then f corresponds to a representation $\pi = \otimes\pi_v$. If $\lambda|\ell$ and $\ell \neq p$ then $\rho_\lambda|D_p$ corresponds to π_p under the Langlands correspondence. The details of this correspondence were figured out by Philip Kutzko, but Carayol completed it in the exceptional case (to be defined later). There are 3 cases to consider.

a) $p\nmid N$

b) $p||N$

c) $p^2|N$

We considered the first two cases last time. The third case is different because the same sort of analysis as we applied to the first two cases no longer works in the sense that we no longer know what π_p looks like.

14.8.1 Possibilities for π_p

Case 1 (principal series) In this case, $\pi_p = \mathrm{PS}(\alpha,\beta)$. Here $\alpha,\beta : \mathbf{Q}_p^* \to \mathbf{C}^*$ are unramified characters. $\mathbf{Q}_p^*$ corresponds to a dense subgroup of the abelian Galois group of $\mathbf{Q}_p$ under the correspondence elucidated in Serre's "Local Classfield Theory" (in Cassels and Frohlich).

$$\begin{array}{ccc} \mathbf{Q}_p^* & \to & \mathrm{Gal}(\overline{\mathbf{Q}_p}/\mathbf{Q}_p)^{\mathrm{ab}} \\ \downarrow & & \downarrow \\ \mathbf{Z} & \to & \hat{\mathbf{Z}} = \mathrm{Gal}(\overline{\mathbf{F}_p}/\mathbf{F}_p) \end{array}$$

Under this correspondence α and β correspond to Galois representations α_λ and β_λ and $\rho_\lambda|D_p \sim \alpha_\lambda \oplus \beta_\lambda$.

Case 2 (special) In this case, π_p is the *special* automorphic representation corresponding to the Galois representation $\kappa \otimes \mathrm{st}$ where st is the Steinberg representation which arises somehow from a split-multiplicative reduction elliptic curve and κ is a Dirichlet character. In this case

$$\rho_\lambda|D_p = \kappa \otimes \begin{pmatrix} \chi_\ell & * \\ 0 & 1 \end{pmatrix}.$$

Case 3 (cuspidal) Case 3 occurs when π_p does not fall into either of the previous cases. Such a π_p is called cuspidal or super-cuspidal. Some of these π_p come from the following recipe. Fix an algebraic closure $\overline{\mathbf{Q}_p}$ of $\mathbf{Q}_p$. Let K be a quadratic extension of $\mathbf{Q}_p$. Let $\psi : K^* \to \mathbf{C}^*$ be a Grössencharacter. Then ψ gives rise to a character

$$\psi_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}_p}/K) \to \overline{E}_\lambda^*$$

which induces $\rho_\lambda|D_p$. That is,

$$\rho_\lambda|D_p = \mathrm{Ind}_{\mathrm{Gal}(\overline{\mathbf{Q}_p}/K)}^{\mathrm{Gal}(\overline{\mathbf{Q}_p}/\mathbf{Q}_p)} \psi_\lambda : \mathrm{Gal}(\overline{\mathbf{Q}_p}/\mathbf{Q}_p) \to \mathrm{GL}(2, \overline{E}_\lambda).$$

This representation is irreducible iff ψ is not invariant under the canonical conjugation of $K/\mathbf{Q}_p$. The pair ψ, K gives rise (via the construction of Jacquet-Langlands) to a representation of $\pi_{\psi,K}$ of $\mathrm{GL}(2, \mathbf{Q}_p)$. Those representations which do not come from this recipe and which do not fall into case 1 or case 2 above are called *extraordinary.* They can only occur when $p \leq 2$.

When can the various cases occur?

Case 1) occurs, e.g, if $p \nmid N$, and also if $p||N$ and ε (the character of f) is ramified at p.

Case 2) occurs if $p||N$ and ε is unramified at p.

14.8.2 The case $\ell = p$

We consider $\rho_\lambda|D_p$ where $\lambda|p$ and $p = \ell$. Write $f = \sum a_n q^n$. The case when $\lambda \nmid a_p$ is called the ordinary case. This case is very similar to the case for an ordinary elliptic curve. In other words,

$$\rho_\lambda|D_p = \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}.$$

[[Ribet mentioned that there is a paper on this by Mazur and Wiles in the American Journal of Mathematics. Look the reference up.]] An important point is that β is unramified so it makes sense to consider $\beta(\mathrm{Frob}_p) = a_p \in E_\lambda^*$. Since $\alpha\beta$ is the determinent, $\alpha\beta = \chi_p^{k-1}\varepsilon = \chi_p\varepsilon$ (after setting $k = 2$ to fix ideas), this gives some description of what is going on. The obvious question to ask is whether or not * is nontrivial. That is, is $\rho_\lambda|D_p$ semisimple or not?

When f has weight 2, then f gives rise to an abelian variety $A = A_f$. Then ρ_λ is defined by looking at the action of Galois on the λ-adic division points on A. If none of the λ lying over p divide a_p then A is ordinary at p. A stronger statement is that the p-divisible group $A[p^\infty]$ has good ordinary reduction.

One simple case is when f has CM. By this we mean that there is a character $\kappa \neq 1$ of order 2 such that $a_n = \kappa(n)a_n$ for all n prime to $\mathrm{cond}(\kappa)$. [[This is not a typo, I do not mean $\overline{a_n} = \kappa(n)a_n$. Coleman said that $a_n = \kappa(n)a_n$ is just a funny way to say that half of the a_n are 0.]] It is easy to prove that f has CM iff the ρ_λ become abelian on some open subgroup of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ of finite index. Ribet explains this in his article in [21]. If f has CM then since the representation ρ_λ is almost abelian one can show that * is trivial. Ribet said he does not know whether the converse is true. Note that f has CM iff $A_f/\overline{\mathbf{Q}}$ has CM. If all $\lambda|p$ are ordinary (i.e., they do not divide a_p) and if * is trivial for every $\rho_\lambda|D_p$ it is easy to show using [26] that A has CM.

Next we will say something more about representations which appear to be ordinary. Consider the situation in which f has weight 2 and p exactly divides the level N of f. Suppose furthermore that the character ε of f is unramified at p. Then π_p is a special representation. The λ-adic representations for $\ell \neq p$ are (up to characters of finite order) like representations attached to some Tate curve. The situation is similar when $\ell = p$ since

$$\rho_\lambda | D_p = \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}.$$

As in the case of a Tate curve $\alpha/\beta = \chi_p$. Up to a quadratic character we know the situation since $\alpha\beta = \chi_p \varepsilon$. Also β is still unramified and $\beta(\mathrm{Frob}_p) = a_p$ is a unit. We know that $a_p^2 = \varepsilon(p)$ is a root of unity. When $k = 2$ the case of a spherical representation mimics what happens for ordinary reduction. The upper right hand entry $*$ is never trivial in this case [[because of something to do with extensions and Kummer theory]].

Ribet said he knows nothing about the situation when $k > 2$. If $p||N$ and ε is unramified at p then π_p is special so ρ_λ, $\ell \neq p$ are again of the form $\left(\begin{smallmatrix} \alpha & * \\ 0 & \beta \end{smallmatrix}\right)$. So we still know everything up to a quadratic character. But if $\ell = p$ some characters are Hodge-Tate [[what does that mean?]] so by a theorem of Faltings they can not be bizarre powers of the cyclotomic character. The multiplicative case like the ordinary case is very special to the case $k = 2$. Wiles uses this heavily in his proof. If k is arbitrary then $a_p^2 = \varepsilon(p) p^{\frac{k}{2}-1}$ so a_p is not a unit for $k > 2$ so there is no invariant line. So the representation is probably irreducible in the case $k > 2$, interestingly enough. [[Echos of "yeah", "strange" and "very strange" are heard throughout the room.]]

14.8.3 Tate Curves

K. Buzzard told me something about Tate curves. Suppose $E/\mathbf{Q}_p$ is an elliptic curve with multiplicative reduction at p and that $j \in \mathbf{Q}_p$ is the j-invariant of E. Using a formula which can be found in Chapter V of [34] one finds $q = q(j)$ with $|q| < 1$. The Tate curve is $E(\overline{\mathbf{Q}}_p) = \overline{\mathbf{Q}}_p / q^{\mathbf{Z}}$. The p torsion on the Tate curve is $\{\zeta_p^a (q^{1/p})^b : 0 \leq a, b \leq p-1\}$. Galois acts by $\zeta_p \mapsto \zeta_p^a$ and $q^{1/p} \mapsto \zeta_p^a q^{1/p}$. Thus the associated Galois representation is $\left(\begin{smallmatrix} \chi_p & * \\ 0 & 1 \end{smallmatrix}\right)$.

Chapter 15

The Weight and Serre's Conjectures

15.1 Introduction

The next topic is a close study of the reduction of these representations mod ℓ. Starting with some eigenform f we make a representation ρ_λ then reducing mod λ gives a representation $\overline{\rho}_\lambda$. The motivating question is

Question 15.1.1. *Which possible eigenforms of various weight and level might give rise to some $\overline{\rho}_\lambda$.*

We might as well assume $\overline{\rho}_\lambda$ is irreducible although interesting work has been done in the reducible case. The main trends in the subject are "lowering" and "raising". The result after much works is that certain of Serre's conjectures are valid in the sense that if ρ_λ arises from a modular form then the level and weight of the modular form is what Serre expected. If ρ_λ "looks like" it comes from a certain space then it does.

The question can also be viewed via the opposite optic. What is the most *bizarre* kind of f that could give rise to ρ_λ? Using certain properties of $\mathrm{cond}(\rho_\lambda)$ one can obtain some sort of control on the possible weights and levels. This is used in the work of Wiles.

15.2 Review of the λ-adic case

A newform f of level N gives rise to a system $(\rho_{\lambda,f})$ of λ-adic representations. We try to describe what goes on with $\rho_\lambda | D_p$. We know everything, in principal at least, when $p \neq \ell$ and $\lambda | \ell$ because of Carayol's theorem. When $p = \ell$ the situation is more complicated. The simplest cases are when

1. $p \nmid N$ and $(a_p, N) = 1$, this is the ordinary good reduction case.

2. $p || N$, the weight of f is 2, and the character ε associated to f is unramified at p.

These two cases are parallel. But if in case 2 the weight is strictly greater than 2 then nothing is known.

15.3 Serre's conjecture 0

Next we talk about the situation for mod ℓ representations. The basic reference is Serre's 1987 article [27].

Conjecture 15.3.1 (Serre's Conjecture 0). Suppose

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F})$$

is a modular representation with $\mathbf{F}$ a finite field. Assume ρ is irreducible and odd, i.e., that $\det(\rho)(c) = -1$ where c is complex conjugation. Then Serre conjectures that ρ is modular.

Formally this means that ρ is isomorphic to a reduction of some $\rho_{f,\lambda}$ where f is a newform on $\Gamma_1(N)$ for some N and λ is a prime of $\mathcal{O} = \mathcal{O}_{E_f}$. Fix an algebraic closure $\overline{\mathbf{F}}$ of $\mathbf{F}$. Then there is an embedding of $\mathcal{O}/\lambda$ into $\overline{\mathbf{F}}$ and a natural inclusion of $\mathbf{F}$ into $\overline{\mathbf{F}}$.

$$\begin{array}{ccc} \mathcal{O}/\lambda & \hookrightarrow & \overline{\mathbf{F}} \\ & & \cup \\ & & \mathbf{F} \end{array}$$

To say that ρ is isomorphic to the reduction of $\rho_{f,\lambda}$ we mean that $\overline{\rho}_{\lambda,f}$ and ρ are isomorphic over $\overline{\mathbf{F}}$.

15.3.1 Problems

One problem is that ρ lives in a slightly smaller field that $\mathcal{O}/\lambda$ would suggest.

Example 15.3.2. Let f be the unique newform [?] of level 23 weight 2 and trivial character. The coefficients of f live in $\mathcal{O} = \mathcal{O}_{\mathbf{Q}(\sqrt{5})} = \mathbf{Z}[\frac{1+\sqrt{5}}{2}]$. For example, a_2 satisfies $x^2 - x - 1 = 0$. Take $\lambda = 2$. Then $\mathcal{O}/\lambda \cong \mathbf{F}_4$. Thus $\overline{\rho}_{\lambda,f}$ lives over $\mathbf{F}_4$. If we just consider those a_p with $p \neq 2, 23$ one can show that these elements all actually live in $\mathbf{Z}[\sqrt{5}]$ which maps mod λ to $\mathbf{F}_2 \subset \mathbf{F}_4$. Because of this it can be shown that $\overline{\rho}_{\lambda,f}$ has a model ρ over $\mathbf{F}_2$. This is because the traces and dets of Frobenius at primes where Frobenius is defined (i.e., $p \neq 2, 23$) take values in $\mathbf{F}_2$. Geometrically, $\mathrm{GL}_2(\mathbf{F}_2) \cong S_3$ so the representation ρ we get is one in which Galois acts as S_3 on 3 points of $X_0(23)$. To see that ρ is modular we might need to pass to an algebraic closure. Starting with $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{F}_2)$ we want to say $\rho \cong \overline{\rho}_{\lambda,f}$. But this isomorphism only takes place over $\mathbf{F}_4$.

At this point K. Buzzard says, "Maybe there is some *better* modular form so that the associated representation actually takes values in $\mathbf{F}_2$ and is isomorphic to ρ. That would be a stronger conjecture." Ribet responds that he has never thought about that question. Next Caskel asks if the determinent of any $\rho_{f,\lambda}$ is a cyclotomic character. The answer is no unless f is a modular form with respect to $\Gamma_0(N)$. Here f is a modular form with respect to $\Gamma_1(N)$.

15.4 Serre's conjecture 1

The next conjecture in Serre's paper [27] is now a theorem when ℓ is odd.

Theorem 15.4.1. *Suppose ℓ is odd. If the mod ℓ representation ρ is irreducible and modular then ρ arises from a newform f of some specific weight $k(\rho)$ and level $N(\rho)$.*

Here $N(\rho)$ is a sort of conductor of ρ defined as before. It is called the "Artin conductor". It is a product $\prod_{p\neq\ell} p^{e(p)}$ where $e(p)$ is a sum

$$\sum_{i=0}^{\infty} \frac{1}{(G_0 : G_i)} \dim(V/V^{G_i}).$$

Here V is the representation space of ρ. The factor $\frac{1}{(G_0:G_i)}$ depends only on $G_0 = \rho|I_p$. [[Should I really say $G_0 = \rho(I_p)$ the image of I_p sitting inside $\mathrm{Aut}(V)$?]] There is, in particular, a term corresponding to $i = 0$ which is $\dim(V/V^{I_p})$. Thus

$$e(p) = \dim(V/V^{I_p}) + \delta(p).$$

Since $\dim V \leq 2$ it follows that $\dim(V/V^{I_p}) \leq 2$. This dimension is 0 if ρ is unramified at p. The term $\delta(p)$ is called the Swan conductor. It is 0 if ρ is tamely ramified at p. There is a relationship between this conductor and something involving f. Suppose $\rho \cong \overline{\rho}_{\lambda,f}$ where f is a newform of level N. It turns out that

$$\mathrm{ord}_p(N) = \dim(V_\lambda/V_\lambda^{I_p}) + \delta(p) = e(p) + \text{ error term}$$

where the error term is $\dim(V^{I_p}) - \dim(V_\lambda^{I_p})$. [[Cryptic statement: The point is that when reducing mod λ more things can become invariant.]] The thing to keep in mind is that if f gives rise to ρ then the power of p in the level of f is constrained since it is given by certain formulas only depending on $e(p)$ and an error term which has magnitude at most 2. The Artin conductor of ρ is some minimal prime to p level which can give rise to ρ. Serre conjectured that there is such an f which gives rise to ρ.

The weight $k(\rho)$ depends only on $\rho|I_\ell$. Thus $k(\rho)$ is some analog for $e(\ell)$. It is kind of hard to know where to begin to explain everything. Why is Serre's conjecture actually a theorem?

15.4.1 Key background points

One idea is that every modular representation ρ must come from an f of level prime to ℓ. Imagine f is a modular form of level ℓ. Consider the corresponding mod ℓ representation. It is unramified outside ℓ so $N(\ell) = 1$. The conjecture then implies this mod ℓ representation comes from a level 1 modular form (i.e., a modular form on $\mathrm{SL}_2(\mathbf{Z})$) of higher weight. This is a classical result.

There is a relationship between mod ℓ forms on $\mathrm{SL}_2(\mathbf{Z})$ of weight $\ell+1$ and mod ℓ forms on $\Gamma_0(\ell)$ of weight 2. The dimensions of these spaces are the same over $\mathbf{F}_\ell$ or over $\mathbf{C}$. There is a map from one space to the other. If $\overline{f}$ is a modular form of weight 2 on $\Gamma_0(\ell)$ multiplication by a certain form gives a form of weight $\ell + 1$ on $\Gamma_0(\ell)$. Then averaging (or taking the trace) gives a modular form $\mathrm{SL}_2(\mathbf{Z})$. The whole process is, in some sense, just the identity map. It was discovered through an appropriate calculation of Serre, 1972, Antwerp. That the formula works was proved (via explicit computations) in the Berkeley Ph.D. thesis of C. Queen [20]. Katz thinks it is easy to prove the formula from the right magical moduli point of view. [[Ribet says he has never seen this.]] Ribet gave a concrete proof for $\ell > 2$ in [22]. It would be nice if someone would construct a clear proof for the case $\ell = 2$. [[I am confused about what Ribet is talking about. What formula?]]

We now allude to the proof of the theorem (Serre's conjecture 1) when $\ell > 2$. Start with a representation ρ. It comes from some (possibly terrible) modular from f.

Step 1. Make the level prime to ℓ via some concrete argument.

Step 2. Try to compare $N(\rho)$ and $N(f)$. These are two prime to ℓ numbers. What happens if $p|\frac{N(f)}{N(\rho)}$? Carayol separated things into appropriate cases. In each case replace f by a better form so the ratio has a smaller power of p in it. This is "level lowering". Eventually we get to the case $N(\rho) = N(f)$. Using a paper of Edixhoven [6] one shows that the weight k of f can be made equal to $k(\rho)$.

What is $k(\rho)$? We give some of the background which leads to Serre's numerical recipe. Start with a modular form f of low weight k. Try to describe what happens if we take the representation attached to f and reduce mod λ then restrict to obtain $\overline{\rho}_{\lambda,f}|I_\ell$. Assume first that we are in the following situation:

- the level $N = N(f)$ is prime to ℓ,
- $\lambda|\ell$,
- the weight k is such that $2 \le k \le \ell + 1$.

There are other cases to consider but we look at this one first. Recall that $\det(\rho) = \chi_\ell^{k-1}\varepsilon$ where $\mathrm{cond}(\varepsilon)|N$. Therefore

$$\det(\rho)|I_\ell = \chi_\ell^{k-1}$$

so we see immediately that $\det(\rho)$ gives k mod $\ell - 1$. This permits the possibility that $k = \ell + 1$ but in fact ρ comes from another form of weight 2, since $2 \equiv \ell + 1 \pmod{\ell - 1}$. But this is only a possibility when 2 different weights could give some ρ. So we should make the following agreement: if there is a form of weight 2 giving ρ then maybe it is better not to look at f. [[There is laughter throughout the room.]]

Next suppose the weight is between 2 and ℓ and we are in the ordinary case so $a_\ell(f) \not\equiv 0 \pmod{\lambda}$. Then $\rho|I_\ell = \begin{pmatrix} \chi_\ell^{k-1} & * \\ 0 & 1 \end{pmatrix}$ exactly as before. [[Ogus says, "* is never 0". Ribet says, "No, no, no. Oh wait, it's my theorem that if this representation were a direct sum of 2 characters, i.e., $* = 0$, then one can find a weight 2 form giving ρ.]] This can be described using Kummer theory but "I do not think about it this way."

15.5 The weight and fundamental characters

Let

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$$

be an irreducible modular mod ℓ representation. Let $D = D_p = \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ be the decomposition group at p. Today we will see how studying the local properties of ρ help us to understand the following theorem.

Theorem 15.5.1. *If $\ell > 2$ and ρ is modular then ρ is modular of a certain weight $k(\rho)$ and level $N(\rho)$.*

Since ρ is modular, $\rho \sim \overline{\rho}_{\lambda,f}$ for some modular form f. The theorem asserts that we can actually take $f \in S_{k(\rho)}(\Gamma_1(N(\rho)))$. As we have seen before $N(\rho)$ is the prime to ℓ Artin conductor of ρ.

Today we attack the question of what $k(\rho)$ should be. To do this we restrict to the inertia group I_p. Before doing this we introduce a certain pair of characters. First we prove a lemma. Let σ be the semisimplification of $\rho|_D$. Thus σ is either a direct sum of two characters or $\rho|_D$ depending on whether or not $\rho|_D$ is irreducible.

Lemma 15.5.2. *The representation σ is tame, i.e, σ is trivial on the ℓ-Sylow subgroup of I_ℓ.*

Proof. Let P be the ℓ-Sylow subgroup of I_ℓ. Since I_ℓ is normal in D and P is unique it follows that P is normal in D. Let $W = \mathbf{F}_{\ell^\nu} \times \mathbf{F}_{\ell^\nu}$ be the representation space of σ. Then

$$W^P = \{w \in W : \sigma(\tau)w = w \quad \text{for all} \quad \tau \in P\}$$

is a subspace of W invariant under the action of D. To see this let $\alpha \in D$ and suppose $w \in W^P$. Since P is normal in D, $\alpha^{-1}\tau\alpha = \tau'$ for some $\tau' \in P$. Therefore

$$\sigma(\alpha)^{-1}\sigma(\tau)\sigma(\alpha)w = \sigma(\tau')w = w$$

so $\sigma(\tau)\sigma(\alpha)w = \sigma(\alpha)w$ hence $\sigma(\alpha)w \in W^P$.

But $W^P \neq 0$. To see this write W as a disjoint union of its orbits under the action of P. Since P is an ℓ-Sylow group and W is finite we see that the size of each orbit is either 1 or a positive power of ℓ. Now $\{0\}$ is a singleton orbit, W has ℓ-power order, and all non-singleton orbits have order a positive power of ℓ so there must be at least $\ell - 1$ other singleton orbits. Each of these other singleton orbits gives a nonzero element of W^P.

If $W^P = W$ then P acts trivially so we are done. If $W^P \neq W$ then W^P is a one dimensional subspace invariant under D so by semisimplicity σ is a diagonal representation. Suppose $\tau \in P$ then τ has order ℓ^n for some n. Write

$$\sigma(\tau) = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

then $\alpha^{\ell^n} = 1$ and $\beta^{\ell^n} = 1$. Since $\alpha, \beta \in \mathbf{F}_{\ell^\nu}$ it follows that $\alpha = \beta = 1$. $\square$

The lemma implies $\sigma|_I$ factors through the tame quotient $I_t = I/P$. We now describe I_t explicitly. Let $\mathbf{Q}_p^{tame}$ be the maximal tame extension of $\mathbf{Q}_p$ and let $K = \mathbf{Q}_p^{ur}$ be the maximal unramified extension. Since P is the part of inertia fixing $\mathbf{Q}_p^{tame}$,

$$I_t = I/P = \mathrm{Gal}(\mathbf{Q}_p^{tame}/\mathbf{Q}_p^{ur}).$$

For each n prime to p there is a tower of fields

$$\begin{array}{c} \mathbf{Q}_p^{tame} \\ | \\ K(p^{\frac{1}{n}}) \\ | \\ K = \mathbf{Q}_p^{ur} \\ | \\ \mathbf{Q}_p \end{array}$$

By Kummer theory $\mathrm{Gal}(K(p^{\frac{1}{n}})/K) = \mu_n(K)$ where $\mu_n(K)$ denotes the group of nth roots of unity in K. Thus for each n prime to p we obtain by restriction a map $I_t \to \mu_n(K)$. They are compatible so passing to the limit we obtain a map

$$I_t \to \varprojlim_{n} \mu_n(K) = \prod_{r \neq p} \varprojlim_{a} \mu_{r^a}(K) = \prod_{r \neq p} \mathbf{Z}_r(1).$$

Viewed more cleverly mod p we obtain a map

$$I_t \to \varprojlim_{n} \mu_n(\overline{\mathbf{F}}_p) = \varprojlim_{n} \mathbf{F}_{p^n}^*.$$

We now fix attention on $n = p^i - 1$. The point is, that for every such i we have a standard map

$$I_t \to \mathbf{F}_{p^i}^*.$$

We call this map the *fundamental character of level i*. This is unnatural because we had to take an embedding $\mathbf{F}_{\ell^\nu} \hookrightarrow \overline{\mathbf{F}}_\ell$. Instead Serre begins with some disembodied field F of order p^i. There are i different maps $\mathbf{F}_{p^i} \to F$ corresponding to the i automorphisms of $\mathbf{F}_{p^i}$. Restricting these maps to $\mathbf{F}_{p^i}^*$ and composing with the fundamental character $I_t \to \mathbf{F}_{p^i}^*$ defined above gives Serre's fundamental characters of I_t with values in F. We have thus defined some characters on I_t which can be viewed as standard.

Recall that we have a mod ℓ^ν representation ρ which gives rise to

$$\sigma : I_t \to \mathrm{GL}(2, \mathbf{F}_{\ell^\nu}).$$

Since the elements of I_t have order prime to the characteristic ℓ this representation is semisimple so it can be diagonalized upon passing to an algebraic closure of $\mathbf{F}_{\ell^\nu}$. Thus σ corresponds to a pair of characters

$$\alpha, \beta : I_t \to \mathbf{F}_{\ell^{2\nu}}^*.$$

These characters have some stability properties since σ is the restriction of a homomorphism from the full decomposition group. Consider the tower of fields

$$K(p^{\frac{1}{n}}) \supset K = \mathbf{Q}_p^{ur} \supset \mathbf{Q}_p.$$

Let $G = \mathrm{Gal}(K(p^{\frac{1}{n}})/\mathbf{Q}_p)$. Recall that $\mathrm{Gal}(K(p^{\frac{1}{n}})/K) = \mu_n(K)$ and $\mathrm{Gal}(K/\mathbf{Q}_p)$ is generated by Frob_p. What happens is that if $h \in \mu_n(K)$ and $g \in G$ is such that g restricts to Frob_p then we have a conjugation formula: $ghg^{-1} = h^p$. Applying this reasoning to the σ situation with $h \in I_t$ and g restricting to Frob_p we find that

$$\sigma(ghg^{-1}) = \sigma(h^p) = \sigma(h)^p.$$

Thus the pair of characters $\{\alpha, \beta\}$ is stable under pth powering, i.e., $\{\alpha, \beta\} = \{\alpha^p, \beta^p\}$. The point is that

$$\sigma(g)\sigma(h)\sigma(g^{-1}) = \sigma(ghg^{-1}) = \sigma(h)^p.$$

Thus the representation $h \mapsto \sigma(h)^p$ is isomorphic to $h \mapsto \sigma(h)$ via conjugation by $\sigma(g)$. Thus α, β as a pair are the same as α^p, β^p as a pair since they came from isomorphic representations.

What does this mean? One possibility is that $\alpha = \alpha^p$ and $\beta = \beta^p$. This means α, β take values in $\mathbf{F}_p^*$. The other possibility is that $\alpha^p = \beta$ and $\beta^p = \alpha$. Then $\alpha^{p^2} = \alpha$ and $\beta^{p^2} = \beta$ so α, β take values in $\mathbf{F}_{p^2}^*$. The first situation in which α and β take values in $\mathbf{F}_p^*$ is called the *level 1* situation. The second situation is called the *level 2* situation.

We will play a carnival game, "guess your weight." First we consider the level 2 case. Our strategy is to try and express α and β in terms of the two fundamental characters of level 2.

First some background. Before stating the general result we talk about a special case. Suppose E is an elliptic curve over $\mathbf{Q}$ and p is a prime (2 is allowed). Assume E has good supersingular reduction at p. Then there is a representation

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(E[p])$$

which may or may not be irreducible. It gives rise via restriction to two characters $\alpha, \beta : I_t \to \mathbf{F}_p^*$. Serre [*Groupes de Galois attachés aux points d'ordre fini des courbes elliptiques sur un corps de nombres*, 1972] proves that α, β are the two fundamental characters of level 1 and that I_t acts tamely and irreducibly. He also obtained a map

$$I_t \to \mathbf{F}_{p^2}^* \subset \mathrm{GL}(2, \mathbf{F}_p)$$

where $\mathbf{F}_{p^2}^*$ sits inside $\mathrm{GL}(2,\mathbf{F}_p)$ via the action of the multiplicative group of a field on itself after choice of a basis. The action of I_t on $\mathbf{F}_{p^2}^*$ is through the fundamental character. This was the world's first view of fundamental characters.

Serre next asked Fontaine the corresponding question for more general weight. The first published proof appears in [6]. Suppose f is a newform of weight k such that $2 \le k \le p$ and that the level of f is prime to $p = \ell$. Take $\overline{\rho}_{f,\lambda}$ where $E_f = \mathbf{Q}(\text{coefficients of } f)$ and λ is a prime of $\mathcal{O}_{E_f}$ dividing p. Suppose we are in the supersingular case so $a_p \equiv 0 \mod \lambda$. Semisimplifying and restricting as before gives a pair of characters $\alpha, \beta : I_t \to \mathbf{F}_{p^\nu}^*$.

Theorem 15.5.3. *Under the above hypothesis, $\{\alpha,\beta\}$ equals $\{\psi^{k-1},(\psi')^{k-1}\}$ where ψ and ψ' are the two fundamental characters of level 2, i.e., $\psi = (\psi')^p$ and $\psi' = \psi^p$.*

Edixhoven's method is to reduce from weight k to weight 2. Given a modular form mod p on $\Gamma_1(N)$ of weight k and character ε he shows that it is enough to look at a corresponding modular form on $\Gamma_1(Np)$ of weight 2 and character $\varepsilon\omega^{k-2}$ where ω is a Teichmuller character on $(\mathbf{Z}/p\mathbf{Z})^*$. This is a method which helps us understand what happens at primes p where a certain abelian variety does not have good reduction but has potentially good reduction. In this direction Ogus suggests reading the paper [Faltings and Jordan, *Crystalline cohomology and* $\mathrm{GL}(2,\mathbf{Q})$, Israel J. Math. 90 (1995), no. 1-3, 1–66.]

What can we guess about the arbitrary situation? The situation is as follows. We begin with a representation ρ which gives rise to a representation σ which in turn gives rise to a pair of fundamental characters ψ^a and $(\psi')^a = \psi^{pa}$. (Here a should be thought of as a number mod p^2-1.) The condition that we are not in level 1 means that a is not divisible by $p+1$ since
$$\psi\psi' = \psi^{p+1} : I_t \to \mathbf{F}_p^*$$
is the unique fundamental character of level 1 (namely the mod p cyclotomic character χ).

We next do a Euclidean division. Take $0 \le a < p^2-1$ and write $a = qp + r$. Then $pa = qp^2 + rp$, but we are working mod p^2-1 so this becomes $pa = q + rp$. What are the possible values for q and r? By the Euclidean algorithm $0 \le r \le p-1$ and $0 \le q \le p-1$. Now what are the constraints on r and q together? The main constraint is that $r \ne q$, since if $r = q$ then a is a multiple of $p+1$. So we can *assume* that $0 \le r < q \le p-1$. We know that
$$\{\alpha,\beta\} = \{(\psi\psi')^r(\psi')^{q-r}, (\psi\psi')^r\psi^{q-r}\}.$$
For example, if $\alpha = \psi^a$ then
$$\alpha = \psi^{qp+r} = (\psi')^q\psi^r = (\psi\psi')^r(\psi')^{q-r}.$$
Since $\psi\psi' = \chi$ we can view $\{\alpha,\beta\}$ as a pair of characters $(\psi')^{q-r}$ and ψ^{q-r} which has been multiplied as a pair by χ^r.

What would happen if for example $r = 0$? In this case
$$\{\alpha,\beta\} = \{(\psi')^{k-1}, \psi^{k-1}\}$$
where $k = q+1$. So when $r = 0$ we guess the weight to be $k(\rho) = q+1$.

What happens more generally? Suppose f is a modular form thought of mod p where $\lambda | p$. If $f = \sum a_n q^n$ gives rise to ρ then we might well ask what modular form gives rise to $\rho \otimes \chi$. In [Lecture Notes in Math, no. 601] we find that $\theta f = \sum n a_n q^n$ is a form giving rise to $\rho \otimes \chi$. If k is the weight of f, then θf has weight $k + p + 1$. More generally, since χ^r appears in the pair
$$\{\psi^{q-r}, (\psi')^{q-r}\} \cdot \chi^r$$

we obtain

$$k(\rho) = q - r + 1 + (p+1)r = q + pr + 1.$$

But be careful, the weight k does not have to go up. In fact, it goes up $p-1$ times then ends up where it started. This gives rise to the theory of theta cycles which we will study in the next lecture.

15.6 The weight in Serre's conjectures on modular representations

Let

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_{p^\nu})$$

be a mod p representation. Let $\ell = p$ and assume that ρ is supersingular and of level 2 at ℓ.

Let ψ and ψ' be the fundamental characters of level 2. Let $I = I_\ell$ be the inertia group at ℓ. After extension of scalars $\rho|_I$ has the form

$$\rho|_I \sim \begin{pmatrix} \psi^a \psi'^b & 0 \\ 0 & \psi'^a \psi^b \end{pmatrix}.$$

We may assume $0 \leq a < b \leq p-1$. Since $\psi' = \psi^p$ we also know that $\psi^a \psi'^b = \psi^{a+pb}$ and $\psi'^a \psi^b = \psi^{b+pa}$.

The conjectured weight is $k(\rho) = b + pa + 1$. Since $\chi = \psi\psi'$ we see that

$$\rho|_I \sim \chi^a \begin{pmatrix} \psi'^{b-a} & 0 \\ 0 & \psi^{b-a} \end{pmatrix}.$$

The weight of $\psi'^{b-a} \oplus \psi^{b-a}$ is $b - a + 1$. Twisting a representation by χ raises the weight by $p+1$ (at least when the weight is in the right range). This is because twisting by χ corresponds to applying the θ operator to the corresponding modular form. Thus

$$k(\rho) = a(p+1) + b - a + 1 = b + pa + 1.$$

15.6.1 θ-series

The theory of the θ operator was first developed by Serre and Swinnerton–Dyer, and later jazzed up by Katz [Lecture Notes in Math, Vol. 601, Springer-Verlag]. There is a notion of modular forms mod p and of q-expansion which gives a map

$$\alpha : \bigoplus_{k \geq 0} M_k(\Gamma_1(N), \mathbf{F}_p) \to \mathbf{F}_p[[q]].$$

This map is far from injective. The kernel is generated by 1 element, a certain Eisenstein series E_p.

Suppose $f \in \mathbf{F}_p[[q]]$ is in the image of α. If $f \neq 0$ let $w(f)$ denote the smallest k so that f comes from some M_k. If f does not coming from any single M_k do not define $w(f)$.

Define an operator θ on $\mathbf{F}_p[[q]]$ by

$$\theta(\sum a_n q^n) = q\frac{d}{dq}(\sum a_n q^n) = \sum n a_n q^n.$$

Work of Serre and Swinnerton-Dyer shows that θ preserves the image of α.

Theorem 15.6.1. *Suppose $f \neq 0$ is a mod p modular form as above. If $w(f) \not\equiv 0 \mod p$ then $w(\theta f) = w(f) + p + 1$.*

The theta series associated to f is

$$w(f), w(\theta f), w(\theta^2 f), \ldots$$

This series is periodic because Fermat's Little Theorem implies $\theta^p f = \theta f$ and so $w(\theta^p f) = w(\theta f)$.

Tate asked the question: What are the possible θ-cycles? That is, what possible sequences $w(f), w(\theta f), \ldots$ can occur? The question was basically answered by work of Edixhoven [Invent. Math. 109, 563-594 (1992)] and Jochnowitz [Trans. A.M.S. vol. 270, 1982 (253-267)].

Suppose f is an eigenform with $2 \leq w(f) \leq p$. Suppose f is supersingular, i.e., $a_p(f) = 0$. Since f is an eigenform the a_n are multiplicative ($a_{nm} = a_n a_m$ for $(n, m) = 1$) and there is a recurrence among prime power coefficients so $a_n(f) = 0$ whenever $p | n$. Thus $\theta^{p-1} f = f$ hence $w(\theta^{p-1} f) = w(f) = k$.

If we start with f and apply θ successively what happens to the weight? First suppose $k = 2$. The θ-cycle is

$$2, 2 + (p + 1), 2 + 2(p + 1), \ldots, 2 + (p - 2)(p + 1).$$

This is because we can apply θ thus raising the weight by $p + 1$ so long as the weight is not a multiple of p. Only the last term in the above sequence is divisible by p. There are $p - 1$ terms so this the full θ-cycle.

Next suppose $k = p$. The θ-cycle is

$$p, 3, 3 + (p + 1), \ldots, 3 + (p + 1)(p - 3).$$

The last term is divisible by p, no earlier term is, and there are $p - 1$ terms so this is the full θ-cycle. We know that the second term must be 3 since it is the only number so that the θ-cycle works out right, i.e., so that the $(p - 1)$st term is divisible by p but no earlier term is. If we would have tried 2 instead of 3 we would have obtained the sequence

$$p, 2, 2 + (p + 1), \ldots, 2 + (p + 1)(p - 3), 2 + (p + 1)(p - 2).$$

This sequence has one too many terms.

Now we consider the general case. Let k be a weight such that $2 < k < p$. The θ-cycle is

$$k, k + (p + 1), \ldots, k + (p + 1)(p - k), k', k' + p + 1, \ldots, k' + (p + 1)(k - 3).$$

The first $p - k + 1$ terms of the sequence are obtained by adding $p + 1$ successively until obtaining a term which is divisible by p. Applying θ to a form of weight $k + (p + 1)(p - k)$ causes the weight to drop to some k'. It can be proved that there is at most one drop in the sequence.

How can we guess k'? It must be such that $k' + (p + 1)(k - 3)$ is divisible by p. Thus the correct answer is

$$k' = p + 3 - k.$$

15.6.2 Edixhoven's paper

Suppose that ρ is an irreducible mod $\ell = p$ representation of level 2. To avoid problems in the exceptional case assume $\ell \geq 3$. Edixhoven proves that if ρ is modular in the sense that ρ comes from an eigenform in $S_k(\Gamma_1(N))$ with $(N,p) = 1$, then ρ comes from an eigenform in $S_{k(\rho)}(\Gamma_1(N))$.

What are the elements of the proof?

1. The behavior of $\rho|_{I_p}$ when $2 \leq w(f) \leq p+1$,
2. the fact that every modular representation ρ coming from a cusp form f has the form $\rho_f \otimes \chi^i$ where $i \in \mathbf{Z}/(p-1)\mathbf{Z}$, and
3. the fact that $w(f)$ satisfies $2 \leq w(f) \leq p+1$.

Serre knew that $2 \leq w(f) \leq p+1$ but never published a proof. How did Serre talk about his result before Edixhoven? The eigenforms f correspond to systems of eigenvalues in $\overline{\mathbf{F}}_p$ of the Hecke operators T_r, $r \not| N$, $r \not| p$. Eigenforms of weight at most $p+1$ give, up to twist, all systems of eigenvalues. A possible proof uses the construction of ρ_f in terms of certain étale cohomology groups.

15.7 The extra assumption

[[The mysterious slogan of the day which Ribet put up this time is "will soon admit".]]

Let $\rho : G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$ be a Galois representation . Assume that ρ is irreducible, $\rho \cong \overline{\rho}_{\lambda,f}$, and $\ell > 2$. Then the theorem is

Theorem 15.7.1. *Under the above hypothesis, ρ comes from a modular form f of weight $k(\rho)$, level $N(\rho)$, and under a certain* extra assumption, *character ε of order prime to ℓ.*

The extra assumption alluded to in the theorem is the following.

Extra Assumption.

Not all of the following are true.

1. $\ell = 3$,
2. $\rho|\,\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}(\sqrt{-3}))$ is abelian, and
3. $\det(\rho)$ is not a power of the mod 3 cyclotomic character χ.

Example 15.7.2. If $\det(\rho) = \chi$ is the mod 3 cyclotomic character, for example when ρ comes from the Galois representation on the 3-torsion of an elliptic curve, then the extra assumption holds so the theorem applies. [[as long as the additional assumption hold?]]

Let ρ be a representation as above. Then

$$\det \rho : G \to \mathbf{F}_{\ell^\nu}$$

is very likely ramified at ℓ. Let χ be the mod ℓ cyclotomic character. As an exercise one can show that $\det \rho = \chi^i \theta$ for some i and some θ where $\theta : G \to \mathbf{F}_{\ell^\nu}$ is unramified at ℓ. Using the Kronecker-Weber theorem and class-field theory one can prove the following two facts.

1. One can write θ as a Dirichlet character $(\mathbf{Z}/M\mathbf{Z})^* \to \mathbf{F}_{\ell^\nu}^*$. [[Does this mean that θ factors through the mod M cyclotomic character?]]

2. In fact $M|N(\rho)$, since the conductor of the determinent of a representation should divide the conductor of the representation. [[?? My notes have the opposite remark?]]

Thus we can view θ as a map $(\mathbf{Z}/N(\rho)\mathbf{Z})^* \to \mathbf{F}_{\ell^\nu}^*$.

Let $H = \ker\theta \subset (\mathbf{Z}/N(\rho)\mathbf{Z})^*$. Define a congruence subgroup $\Gamma_H(N)$ by

$$\Gamma_H(N) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) : a, d \in H\}.$$

We have the following theorem.

Theorem 15.7.3. *Suppose $\ell > 2$ and ρ satisfies the above assumptions including the extra assumption. Then ρ arises from a form f in $S_{k(\rho)}(\Gamma_H(N(\rho)))$.*

In particular, the theorem is true if ρ comes from the ℓ-torsion representation on an elliptic curve since $\det = \chi$ so the extra assumption is satisfied.

One might ask if the extra assumption is really necessary. At first Serre did not think that it was. But he was surprised to discover the following example which shows that the extra assumption can not be eliminated.

Example 15.7.4. (Serre) The space $S_2(\Gamma_1(13))$ is 2 dimensional. Let $\ell = 3$. There is a newform $f \in S_2(\Gamma_1(13))$ of character $\varepsilon : (\mathbf{Z}/13\mathbf{Z})^* \to \mathbf{C}^*$. This character must be trivial on $\{\pm 1\}$ and has order 6. Using [[cite Ribet's LNM 601 article]] one shows that

$$\mathbf{Q}(\ldots, a_n, \ldots) \supset \mathbf{Q}(\ldots, \varepsilon(d), \ldots) = \mathbf{Q}(\sqrt{-3}).$$

Furthermore, since $S_2(\Gamma_1(13))$ has dimension 2 there can be at most 2 embeddings of $\mathbf{Q}(\ldots, a_n, \ldots)$ into $\mathbf{C}$ (each embedding gives an independent newform). Thus one actually has that $\mathbf{Q}(\sqrt{-3}) = \mathbf{Q}(\ldots, a_n, \ldots)$.

Let $\lambda = \sqrt{3}$ and let $\overline{\rho}_{\lambda,f} : G \to \mathrm{GL}(2, \mathbf{F}_3)$ be the associated Galois representation . Then $\det(\overline{\rho}_{\lambda,f}) = \chi\theta$ where χ is the mod 3 cyclotomic character and $\theta \equiv \varepsilon \pmod 3$ [[I am not really sure what this congruence means]]. In particular, θ has order 2. Thus $H = \ker(\theta) \subset (\mathbf{Z}/13\mathbf{Z})^*$ is exactly the squares in $(\mathbf{Z}/13\mathbf{Z})^*$.

The conclusion of the theorem can not hold since $S_2(\Gamma_H(13)) = 0$. This is because any form would have to have a character whose order is at most two since it must be trivial on H, but $S_2(\Gamma_H(13)) \subset S_2(\Gamma_1(13))$ and $S_2(\Gamma_1(13))$ consists of forms whose character is of order 6. [[There are a lot of compatibilities to check here which I have not checked.]]

In this example

- $\ell = 3$,
- $\det\rho$ is not a power of χ, and
- $\overline{\rho}_{\lambda,f} |\, \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}(\sqrt{-3}))$ is abelian.

A good way to see the last assertion is to look at the following formula:

$$f \otimes \varepsilon^{-1} = \overline{f}$$

(up to an Euler factor at 13) in the sense that

$$\rho_{f,\lambda} \otimes \varepsilon^{-1} = \rho_{\overline{f}}.$$

Now reduce mod 3 to obtain

$$\overline{\rho}_{\lambda,f} \otimes \varepsilon^{-1} \cong \overline{\rho}_{\lambda,f}$$

since $f \equiv \overline{f} \pmod{\sqrt{-3}}$ (since 3 is ramified). Thus $\overline{\rho}_{\lambda,f}$ is isomorphic to a twist of itself by a complex character thus by elementary facts $\overline{\rho}_{\lambda,f}$ is reducible and abelian over the field corresponding to its kernel. [[So why is the Galois group of $\mathbf{Q}(\sqrt{-3})$ in the kernel of $\overline{\rho}_{\lambda,f}$?]] [[Ribet suddenly gets very excited and notes that $\overline{\rho}_{\lambda,f}$ is also abelian on the Galois groups of $\mathbf{Q}(\sqrt{13})$ and $\mathbf{Q}(\sqrt{39})$ by the same argument.]]

There are two main ingredients in this theorem. The first is companion forms.

15.7.1 Companion Forms

The key papers dealing with companion forms are Gross and Coleman-Voloch. The context is the following. Suppose we are interested in the recipe for $k(\rho)$. There are two cases to consider, the level 1 case and the level 2 case. We have already discussed the level 2 case. In the level 1 case what happens? If we semisimplify and restrict to inertia we obtain a direct sum of two representations. There are two possibilities for $\rho|I$:

$$\rho|I = \begin{pmatrix} \chi^\alpha & * \\ 0 & \chi^\beta \end{pmatrix} \text{ or } \rho|I = \begin{pmatrix} \chi^\alpha & 0 \\ 0 & \chi^\beta \end{pmatrix}.$$

In the second case it is easy to guess $k(\rho)$. The thing to do is to look at the exponents and normalize as best you can. Since α and β are only defined mod $p-1$ we may, after relabeling if necessary, assume that $0 \le \alpha \le \beta \le p-2$. Factor out χ^α to obtain

$$\chi^\alpha \otimes \begin{pmatrix} 1 & 0 \\ 0 & \chi^{\beta-\alpha} \end{pmatrix}.$$

Next (secretly) recall that if f is ordinary of weight k then f gives rise to the representation

$$\begin{pmatrix} \chi^{k-1} & * \\ 0 & \chi \end{pmatrix}$$

(here $*$ can be trivial). [[Something must be wrong here because nothing makes sense in the following argument unless the lower right entry of this matrix is 1 instead of χ.]] If f is of weight $\beta - \alpha + 1$ twisting α times by $q\frac{d}{dq}$ gives the desired representation. Thus the weight $k(\rho)$ is

$$(p+1)\alpha + \beta - \alpha + 1 = \beta + p\alpha + 1.$$

There is one caveat. Since Serre was terrified of weight 1 forms, if $\alpha = \beta = 0$ then Serre defines $k(\rho) = p$ instead of $k(\rho) = 1$. There is no distinction between Serre's formula and the above formula when the weight is > 1 since, as Katz proved, any form of weight > 1 lifts to a characteristic 0 form. But this is not true for weight 1. To understand the weight 1 forms we must use sophisticated tools of algebraic geometry.

Suppose f is of weight k, $2 \le k \le p+1$. Let $\overline{\rho}_{\lambda,f}$ be the associated representation where λ is some prime dividing $\ell = p$. Then

$$\overline{\rho}_{\lambda,f}|I_p = \begin{pmatrix} \chi^{k-1} & * \\ 0 & 1 \end{pmatrix}.$$

[[I really wonder why this should be so in such brave generality? It seems like things were not like this just a minute ago.]] To introduce the idea of a companion form suppose that somehow by chance $* = 0$ so we have $\begin{pmatrix} \chi^{k-1} & 0 \\ 0 & 1 \end{pmatrix}$. Take ρ and twist by χ^{1-k} then

$$\rho \otimes \chi^{1-k}|I_p = \rho \otimes \chi^{p-k}|I_p = \begin{pmatrix} 1 & 0 \\ 0 & \chi^{p-k} \end{pmatrix}.$$

What is the minimum weight giving rise to this representation. The level is 1 [[why??]] the representation is semisimple and reducible and $\alpha = 0$, $\beta = p - k$, so the natural weight is $p - k + 1$. Thus the conjecture predicts that there should exist another form g of weight $p+1-k$ such that $\rho_g = \rho_f \otimes \chi^{p-k}$. Here g is a *companion form* of f. Thus $g = \theta^{p-k} f$ where $\theta = q\frac{d}{dq}$ is the θ-operator we defined when studying θ-series. This conjecture was proved by Gross when $k \neq 2, p$. It was proved by Coleman and Voloch when $k = p$. This is how we know about the weight. Next time we will talk about the character.

[[What is the other main ingredient in the proof.]]

15.8 The exceptional level 1 case

Let

$$\rho : G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$$

be a Galois representation with $\ell > 2$. Assume that ρ is irreducible and modular. Then ρ comes from a modular form in $S_{k(\rho)}(\Gamma_1(Np))$. Here $p = \ell$ [I think] and $2 \leq k(\rho) \leq p^2 - 1$.

We still must talk about the exceptional case in level 1. In this situation

$$\rho|I_\ell = \begin{pmatrix} \chi^\alpha & * \\ 0 & \chi^\beta \end{pmatrix} = \chi^\beta \otimes \begin{pmatrix} \chi & * \\ 0 & 1 \end{pmatrix}$$

since $\alpha = \beta + 1$. [[I do not remember why.]] Define a representation σ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ by

$$\sigma := \rho \otimes \chi^{-\beta} \cong \begin{pmatrix} \chi & * \\ 0 & 1 \end{pmatrix}.$$

Fortunately we know enough to say something about the weight $k(\sigma)$. In fact,

$$k(\sigma) = \begin{cases} 2 & \sigma \text{ if is finite at } \ell \\ \ell + 1 & \text{otherwise} \end{cases}$$

This is enough to determine $k(\rho)$ since

$$k(\rho) = k(\chi^\beta \otimes \sigma) = (p+1)\beta + k(\sigma).$$

We have not yet said what we mean for σ to be finite at ℓ.

Definition 15.8.1. $\sigma|D_\ell$ is finite if $\sigma|D_\ell$ is equivalent to a representation defined by $\mathcal{G}(\overline{\mathbf{Q}}_\ell)$, where $\mathcal{G}$ is a finite flat $\mathbf{F}_{\ell^\nu}$ vector space scheme over $\mathbf{Z}_\ell$.

This definition is not terribly enlightening. To help fathom what is going on we consider the following special case. Suppose $E/\mathbf{Q}$ is an elliptic curve with semistable (=good or multiplicative) reduction at ℓ. Then $E[\ell]$ defines a representation

$$\sigma : G \to \mathrm{Aut}\, E[\ell] = \mathrm{GL}(2, \mathbf{F}_\ell).$$

Let Δ_E be the discriminant of E, then σ is finite at ℓ iff

$$\mathrm{ord}_\ell \Delta_E \equiv 0 \pmod{\ell}.$$

Thus when Δ_E satisfies this requirement, $k(\sigma) = 2$. Next suppose $p \neq \ell$ and assume E has semistable reduction at p. Then the representation σ attached to $E[\ell]$ is unramified at p iff

$$\mathrm{ord}_p \Delta_E \equiv 0 \pmod{\ell}.$$

The reference for everything we have been discussing is [27].

The bridge is Kummer theory. Let E be an elliptic curve with multiplicative reduction at p. Suppose $V = E[\ell] = \mathbf{F}_\ell \oplus \mathbf{F}_\ell$. We have a map

$$\tau : D = \mathrm{Gal}(\overline{\mathbf{Q}}_\ell/\mathbf{Q}_\ell) \to \mathrm{Aut}\, V.$$

We treat the case $p = \ell$ and $p \neq \ell$ at the same time. By the theory of Tate curves there is an exact sequence of D-modules

$$0 \to X \to V \xrightarrow{\alpha} Y \to 0.$$

[[I do not remember him explicitly saying that these maps should be D-modules maps but they must be for the twist ${}^d s$ to be a splitting.]] Each of these terms is a D module and X and Y are 1 dimensional. The action of D on Y is given by an unramified character ε of order dividing 2 (i.e., the order is 1 or 2). The action of D on X is given by $\chi\varepsilon$.

Next we define an element of $H^1(D, \mathrm{Hom}(Y, X))$. Choose a splitting $s : Y \to V$. Thus s is a linear map (not necessarily a map of D-modules) [[otherwise all twists of a splittings would be trivial]] such that $\alpha s = 1$. For each $d \in D$ consider the twisting ${}^d s : Y \to V$ defined by ${}^d s(y) = ds(d^{-1}y)$. Since

$$\alpha(ds(d^{-1}y)) = d(\alpha(s(d^{-1}y))) = d(d^{-1}y) = y$$

if follows that ${}^d s$ is again a splitting. Thus ${}^d s - s : Y \to V$ followed by $\alpha : V \to Y$ is 0. We therefore have that ${}^d s - s \in \mathrm{Hom}(Y, X)$. The map $d \mapsto {}^d s$ defines an element of $H^1(D, \mathrm{Hom}(Y, X))$.

We have isomorphisms

$$H^1(D, \mathrm{Hom}(Y, X)) \cong H^1(D, \mu_\ell) \cong \mathbf{Q}_p^*/\mathbf{Q}_p^{*\ell}.$$

The last isomorphism follows from Kummer theory. [[I do not know where the first one comes from.]] Thus τ defines an element of $\mathbf{Q}_p^*/\mathbf{Q}_p^{*\ell}$. Serre proved that τ is finite iff the corresponding element of $\mathbf{Q}_p^*/\mathbf{Q}_p^{*\ell}$ defined by τ is in the image of $\mathbf{Z}_p^*$.

If $E/\mathbf{Q}_p$ is an elliptic curve with multiplicative reduction then there exists a Tate parameter $q \in \mathbf{Q}_p^*$ with $\mathrm{Val}_p(q) > 0$ such that

$$E \cong E_q := \mathbf{G}_m/q^{\mathbf{Z}}$$

over the unique quadratic unramified extension of $\mathbf{Q}_p$. The kernel of multiplication by p gives rise to an exact sequence as above. [[Work out the details sometime!]] Furthermore, the element of $\mathbf{Q}_p^*/\mathbf{Q}_p^{*^\ell}$ defined by the representation comming from $E[\ell]$ is just the image of q. In this context there is a formula for Δ_E which is

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

Note that the product factor is a unit in $\mathbf{Q}_p$. Thus $\mathrm{Val}_p \Delta_E = \mathrm{Val}_p q$.

[[At this moment Ogus asks if Serre uses Raynaud's result. Ribet first says no, but then changes his mind and says "Yes, he must be able to explicitly describe finite flat vector space schemes." Then Ogus asks what is wrong with weight 1. Ribet replies that Serre didn't know how to define forms in weight 1. But L. Merel says something about Serre being frustrated because he couldn't compute in weight 1.]]

Chapter 16

Fermat's Last Theorem

16.1 The application to Fermat

"As part of this parcel, I can sketch the application to Fermat."

Suppose that semistable elliptic curves over $\mathbf{Q}$ are modular. Then FLT is true. Why? "As I have explained *so many times* ... " Suppose $\ell > 5$ and

$$a^\ell + b^\ell + c^\ell = 0$$

with $abc \neq 0$, all relatively prime, and such that $A = a^\ell \equiv -1 \pmod 4$, $B = b^\ell$ is even and $C = c^\ell \equiv 1 \pmod 4$. Then we consider the elliptic curve

$$E : y^2 = x(x - A)(x - B).$$

The *minimal discriminant* is

$$\Delta_E = \frac{(ABC)^2}{2^8}$$

as discussed in Serre's [27] and Diamond-Kramer [[cite Math Research Letters, 1995]]. The conductor N_E is equal to the product of the primes dividing ABC (so in particular N_E is square-free). Furthermore, E is semistable – the only hard place to check is at 2. Diamond-Kramer checks this explicitly.

Here is how to get Fermat's theorem. View $E[\ell]$ as a $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-module. The idea is to show that this representation must come from a modular form of weight 2 and level 2. This will be a contradiction since there are no modular forms of weight 2 and level 2. But to apply the level and weight theorem we need to know that $E[\ell]$ is irreducible. The proof of this is due to Mazur.

Let $\rho : G \to \operatorname{Aut} E[\ell]$ be the Galois representation on the ℓ torsion of E. Since E is semistable for $p \neq \ell$,

$$\rho|I_p \cong \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}.$$

Assume ρ is reducible. Then ρ has an invariant subspace so

$$\rho \cong \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}.$$

Then from the form of $\rho|I_p$ we see that the characters α and β could be ramified only at ℓ. Thus $\alpha = \chi^i$ and $\beta = \chi^{1-i}$ where χ is the mod ℓ cyclotomic character. The exponents are i and $1-i$ since $\alpha\beta$ is the determinent which is χ. [[Why is χ supposed to be the only possible unramified character? probably since whatever the character is, it is a product of χ times something else, and the other factor is ramified.]]

What happens to ρ at ℓ, i.e., what is $\rho|I_\ell$? There are only two possibilities. Either $\rho|I_\ell$ is the direct sum of the two fundamental characters or it is the sum of the trivial character with χ. The second possibility must be the one which occurs. [[I do not understand why... something about "characters globally are determined by local information."]] So either $i=0$ or $i=1$. If $i=0$,

$$\rho = \begin{pmatrix} 1 & * \\ 0 & \chi \end{pmatrix}.$$

This means that there is an element of $E[\ell]$ whose subspace is left invariant under the action of Galois. Thus E has a point of order ℓ rational over $\mathbf{Q}$. If $i=1$ then

$$\rho = \begin{pmatrix} \chi & * \\ 0 & 1 \end{pmatrix}.$$

Therefore $\mu_\ell \hookrightarrow E[\ell]$. Divide to obtain $E' = E/\mu_\ell$. The representation on $E[\ell]/\mu_\ell$ is constant (this is basic linear algebra) so the resulting representation on $E'[\ell]$ has the form $\left(\begin{smallmatrix} 1 & * \\ 0 & \chi \end{smallmatrix}\right)$ so E' has a rational point of order ℓ.

Now $E[2]$ is a trivial Galois module since it contains 3 obvious rational points, namely $(0,0), (A,0)$, and $(B,0)$. Thus the group structure on the curve E (or E') (which we constructed from a counterexample to Fermat) contains

$$\mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/\ell\mathbf{Z}.$$

In Mazur's paper [["rational isogenies of prime degree"]] he proves that $\ell \leq 3$. Since we assumed that $\ell > 5$, this is a contradiction. Notice that we have *not* just proved FLT. We have demonstrated the *irreducibility* of the Galois representation on the ℓ torsion of the elliptic curve E arising from a hypothetical counterexample to FLT.

We now have a representation

$$\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_\ell) = \mathrm{Aut}\, E[\ell]$$

which is irreducible and modular of weight 2 and level $N = N_E$ (the conductor of E). Because ρ is irreducible we conclude that ρ is modular of weight $k(\rho)$ and level $N(\rho)$. Furthermore

$$\mathrm{ord}_\ell\, \Delta_E = \mathrm{ord}_\ell (ABC)^2 = 2\ell\, \mathrm{ord}_\ell\, abc \equiv 0 \pmod{\ell}$$

so $k(\rho) = 2$.

We can also prove that $N(\rho) = 2$. Clearly $N(\rho)|N_E$. This is because $N(\rho)$ computed locally at $p \neq \ell$ divides the power of p in the conductor of the ℓ-adic representation for E at p. [[I do not understand this.]] Since ρ is only ramified at 2 or maybe ℓ, $N(\rho)$ must be a power of 2. For $p \neq \ell$, ρ is ramified at p iff $\mathrm{ord}_p\, \Delta_E \not\equiv 0 \pmod{\ell}$ which does happen when $p = 2$. Since N_E is square free this implies that $N(\rho) = 2$. But $S_2(\Gamma_1(2)) = 0$, which is the ultimate contradiction!

But how do we know semistable elliptic curves over $\mathbf{Q}$ are modular?

16.2 Modular Elliptic Curves

Theorem 16.2.1 (Theorem A). *Every semistable elliptic curve over* $\mathbf{Q}$ *is modular.*

There are several ways to define a modular elliptic curve.

Definition 16.2.2. Let E be an elliptic curve. Then E is *modular* if there is a prime $\ell > 2$ such that the associated ℓ-adic Galois representation

$$\rho_{E,\ell^\infty} : G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{Z}_\ell)$$

defined by the ℓ-power division points on E is modular (i.e., it is a $\rho_{f,\lambda}$).

Definition 16.2.3. Let E be an elliptic curve of conductor N_E. For each prime p not dividing N_E one associates a number a_p related [[in a simple way!]] to the number of points on the reduction of E modulo p. Then E is *modular* if there exists a cusp form $f = \sum b_n q^n$ which is an eigenform

for the Hecke operators such that $a_p = b_p$ for almost all p. In the end one deduces that f can be chosen to have weight 2, trivial character, and level N_E.

Definition 16.2.4 (Shimura). An elliptic curve E is modular if there is nonconstant map $X_0(N) \to E$ for some N.

Theorem 16.2.5 (Theorem B. Wiles, Taylor-Wiles). *Suppose E is a semistable elliptic curve over $\mathbf{Q}$ and suppose ℓ is an odd prime such that $E[\ell]$ is irreducible and modular, then ρ_{E,ℓ^∞} is modular and hence E is modular.*

Now we sketch a proof that theorem B implies theorem A. First take $\ell = 3$. If $E[3]$ is irreducible then by work of Langlands-Tunnel we win. The idea is to take

$$\rho : G \to \mathrm{GL}(2, \mathbf{F}_3) \hookrightarrow \mathrm{GL}(2, \mathbf{Z}[\sqrt{-2}]) \subset \mathrm{GL}(2, \mathbf{C}).$$

The point is that there are two maps $\mathbf{Z}[\sqrt{-2}] \to \mathbf{F}_3$ given by reduction modulo each of the primes lying over 3. Choosing one gives a map $\mathrm{GL}(2, \mathbf{Z}[\sqrt{-2}]) \to \mathrm{GL}(2, \mathbf{F}_3)$ which, for some amazing reason [[which is?]], has a section. Then $\rho : G \to \mathrm{GL}(2, \mathbf{C})$ is a continuous representation with odd determinent which must still be irreducible. By Langlands-Tunnel we know that ρ is modular and in fact ρ comes from a weight 1 cusp form f of level a power of 3 times powers of most primes dividing $\mathrm{cond}(E)$. Reducing f modulo some prime of $\mathbf{Z}[\sqrt{-2}]$ lieing over 3 we obtain a mod 3 modular form which corresponds to $\rho : G \to E[3]$. The proof of all this uses the immense base-change business in Langlands' book. [[Ribet next says: "have to get 3's out of the level! This jacks up the weight, and the level is still not square free. Then have to adjust the weight again." I do not know what the point of this is.]]

Kevin Buzzard asked a question relating to how one knows the hypothesis needed for the theorem on weights and levels applies in our situation. To answer this suppose $\ell = 3$ or 5. Form the associated representation $\rho : G \to \mathrm{GL}(2, \mathbf{F}_\ell)$ coming from $E[\ell]$ and assume it is irreducible, modular and semistable.

Definition 16.2.6. A mod ℓ Galois representation is *semistable* if for all $p \neq \ell$, the inertia group I_p acts unipotently and the conjectured weight is 2 or $\ell + 1$.

Note that $\det \rho = \chi$.

Lemma 16.2.7. *Under the above assumptions ℓ divides the order of the image of ρ.*

Proof. If not, then ρ is finite at all primes p, since for primes $p \neq \ell$ inertia acts trivially [[some other argument for ℓ]]. Inertia acts trivially because if the order of the image of ρ is prime to ℓ then ρ acts diagonally. For if not then since $\rho|I_p$ is unipotent (hence all eigenvalues are 1), in a suitable basis something like $\left(\begin{smallmatrix} 1 & \psi \\ 0 & 1 \end{smallmatrix}\right)$ is in the image of ρ and has order ℓ, a contradiction. Because of finiteness $k(\rho) = 2$ and $N(\rho) = 1$ which is a contradiction since there are no forms of weight 2 and level 1. $\square$

Next we consider what happens if $E[3]$ is reducible. There are two cases to consider. First suppose $E[5]$ is also reducible. Then E contains a rational subgroup of order 15. We can check by hand that all such curves are modular. The key result that is that $X_0(15)(\mathbf{Q})$ is finite.

The second possibility is that $E[5]$ is irreducible. In this case we first find a curve E' which is semistable over $\mathbf{Q}$ such that

- $E'[5] \cong E[5]$ (this is easy to do because of the lucky coincidence that $X_0(5)$ has genus 0)
- $E'[3]$ is irreducible

Next we discover that E' is modular since $E'[3]$ is irreducible. This implies $E'[5]$ is modular hence $E[5]$ is modular. Theorem B then implies E is modular.

Chapter 17

Deformations

17.1 Introduction

For the rest of the semester let ℓ be an odd prime. Let $\rho : G \to \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$ be such that

- ρ is modular
- ρ is irreducible
- ρ is semistable

Definition 17.1.1. The representation ρ is *semistable* if

- for all $p \neq \ell$,

$$\rho|I_p \cong \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix},$$

  ($*$ is typically trivial since most primes are unramified.)

- $k(\rho) = 2$ or $\ell + 1$.

This means that there are 2 possibilities.

1. ρ is finite at D_ℓ.
2. $$\rho|D_\ell \cong \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}$$

   where β is unramified. (Since $\det(\rho) = \chi$ we can add that $\alpha|I_\ell = \chi$.)

If $k(\rho) = 2$ then possibility 1 occurs. If $k(\rho) = \ell + 1$ then we are in case 2, but being in case 2 does not imply that $k(\rho) = \ell + 1$. If ρ comes from an elliptic curve $E/\mathbf{Q}$, then case 1 occurs if E has good reduction at ℓ whereas case 2 occurs if E has ordinary multiplicative reduction [[I am not sure about this last assertion because I missed it in class.]].

What is a deformation of ρ and when can we prove that it is modular?

Let A be a complete local Noetherian ring with maximal ideal $\mathbf{m}$ and residue field $\mathbf{F}_{\ell^\nu}$ (so A is furnished with a map $A/\mathbf{m} \xrightarrow{\sim} \mathbf{F}_{\ell^\nu}$). Let

$$\tilde{\rho}_A : G \to \mathrm{GL}(2, A)$$

be a representation which is ramified at only finitely many primes. Assume $\tilde{\rho} = \tilde{\rho}_A$ lifts ρ, i.e., the reduction of $\tilde{\rho}$ mod $\mathbf{m}$ gives ρ.

Theorem 17.1.2. *Let the notation be as above, then $\tilde{\rho}$ is modular iff it satisfies ($*$).*

Neither of the terms in this theorem have been defined yet.

17.2 Condition (∗)

Let

$$\rho : G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$$

be a Galois representation.

The statement we wish to understand is

"All $\tilde{\rho}$ which satisfy (∗) are modular."

Let A be a complete local Noetherian ring with residue field $\mathbf{F}_{\ell^\nu}$. This means that we are given a map $A/\mathbf{m} \cong \mathbf{F}_{\ell^\nu}$. Suppose $\tilde{\rho} : G \to \mathrm{GL}(2, A)$ satisfies the following conditions:

- $\tilde{\rho}$ lifts ρ,
- $\det \tilde{\rho} = \tilde{\chi}$,
- $\tilde{\rho}$ is ramified at only finitely many primes, and
- condition (∗).

What is condition (∗)? It the requirement that $\tilde{\rho}$ have the same qualitative properties as ρ locally at ℓ. There are two cases to consider.

Case 1. (arising from supersingular reduction at ℓ) Suppose ρ is finite and flat at ℓ. Then $\rho|I_\ell$ is given by the 2 fundamental characters

$$I_\ell \to \mathbf{F}_{\ell^2}^*$$

of level 2 (instead of from powers of these characters because of the semistability assumption). Condition (∗) is that the lift of ρ is also constrained to be finite and flat. This means that for every $n \geq 1$,

$$\tilde{\rho}|D_\ell \mod \mathbf{m}^n : D_\ell \to \mathrm{GL}(2, A/\mathbf{m}^n)$$

is finite and flat, i.e., it comes from a finite flat group scheme over $\mathbf{Z}_\ell$ which is provided with an action of $A/\mathbf{m}^n$ as endomorphisms.

Case 2. (bad multiplicative reduction at ℓ or good ordinary reduction at ℓ) In case 2

$$\rho|D_\ell \cong \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}$$

where β is unramified (equivalently $\alpha|I_\ell = \chi$). [[In the case of good ordinary reduction β is given by the action of Galois on $E[\ell]$ in characteristic ℓ.]] Condition (∗) is the requirement that

$$\tilde{\rho}|D_\ell \cong \begin{pmatrix} \tilde{\alpha} & * \\ 0 & \tilde{\beta} \end{pmatrix},$$

where $\tilde{\beta}$ is unramified, $\tilde{\alpha}|I_\ell = \tilde{\chi}$. It follows automatically that $\tilde{\beta}$ lifts β.

17.2.1 Finite flat representations

In general what does it mean for ρ to be finite and flat. It means that there exists a finite flat group scheme $\mathcal{G}$ over $\mathbf{Z}_\ell$ such that $\mathcal{G}(\overline{\mathbf{Q}}_\ell)$ is the representation space of ρ. This definition is *subtle.*

Coleman asked if there is a way to reformulate the definition without mentioning group schemes. Ribet mentioned Hopf algebras but then stopped. Buzzard suggested some of the subtlety of the definition by claiming that in some situation χ is finite flat whereas χ^2 is not. Ogus vaguely conjectured that Fontaine's language is the way to understand this.

It is possible in case 2 above for $\rho|D_\ell$ to be finite flat without $\tilde{\rho}$ finite flat. The quintessential example is an elliptic curve E with supersingular reduction at ℓ such that $\ell|\operatorname{ord}_\ell \Delta_E$.

17.3 Classes of Liftings

Let Σ be a finite set of prime numbers. We characterize a class of liftings $\tilde{\rho}$ which depends on Σ. What does it mean for $\tilde{\rho}$ to be in the class of deformations corresponding to Σ?

17.3.1 The case $p \neq \ell$

First we talk about the case when $p \neq \ell$. If $p \in \Sigma$ then there is no special condition on $\tilde{\rho}|I_p$. If $p \notin \Sigma$ one requires that $\tilde{\rho}$ is qualitatively the same as ρ. This means

1. If ρ is unramified at p (which is the usual case), then we just require that $\tilde{\rho}$ is unramified at p.

2. If ρ is ramified but unipotent at p so

$$\rho|I_p \cong \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$$

we require that

$$\tilde{\rho}|I_p \cong \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}.$$

This situation can occur with an elliptic curve which has multiplicative reduction at p and for which $\ell \nmid \operatorname{ord}_p \Delta_E$.

At this point I mention the prime example.

Example 17.3.1. Suppose $\tilde{\rho} = \rho_{f,\lambda}$ is the λ-adic representation attached to f. What can we say about $\operatorname{ord}_p N(f)$? We know that

$$\operatorname{ord}_p N(f) = \operatorname{ord}_p N(\rho) + \dim(\rho)^{I_p} - \dim(\tilde{\rho})^{I_p}$$

where $(\rho)^{I_p}$ means the inertia invariants in the representation space of ρ. If ρ is semistable then

$$\operatorname{ord}_p N(\rho) + \dim(\rho)^{I_p} = 2.$$

Since we are assuming ρ is semistable, $\operatorname{ord}_p N(f) \leq 2$. Furthermore, the condition $p \notin \Sigma$ is a way of saying

$$\operatorname{ord}_p N(f) = \operatorname{ord}_p N(\rho).$$

Thus the requirement that $p \notin \Sigma$ is that the error term $\dim(\rho)^{I_p} - \dim(\tilde{\rho})^{I_p}$ vanish.

Note that $\operatorname{ord}_p N(\tilde{\rho}) = \operatorname{ord}_p N(f)$ by Carayol's theorem. Thus $\operatorname{ord}_p N(f)$ is just a different way to write $\operatorname{ord}_p N(\tilde{\rho})$.

Example 17.3.2. Imagine ρ is ramified at p and $\operatorname{ord}_p N(\rho) = 1$. Then $\operatorname{ord}_p N(f)$ is either 1 or 2. The requirement that $p \notin \Sigma$ is that $\operatorname{ord}_p N(f) = 1$.

17.3.2 The case $p = \ell$

Next we talk about the case $p = \ell$. There are two possibilities: either $\ell \in \Sigma$ or $\ell \notin \Sigma$. If $\ell \in \Sigma$ then we impose no *further* condition on $\tilde{\rho}$ (besides the already imposed condition $(*)$, semistability at ℓ, etc.). If $\ell \notin \Sigma$ and ρ if finite and flat (which is not always the case) then we require $\tilde{\rho}$ to be finite and flat. If $\ell \notin \Sigma$ and ρ is not finite flat then no further restriction (this is the Tate curve situation).

Suppose $\tilde{\rho} = \rho_{f,\lambda}$, and $\tilde{\rho}$ belongs to the class defined by Σ. We want to *guess* (since there is no Carayol theorem) an integer N_Σ such that $N(f)|N_\Sigma$. What is N_Σ? It will be

$$N_\Sigma = \prod_{\substack{p\neq\ell \\ p\in\Sigma}} p^2 \cdot \prod_{\substack{p\neq\ell \\ p\notin\Sigma}} p^{\operatorname{ord}_p N(\rho)} \cdot \ell^\delta.$$

Here $\operatorname{ord}_p N(\rho)$ is 1 iff ρ is ramified at p. The exponent δ is either 0 or 1. It is 1 iff $k(\rho) = \ell+1$ or $\ell \in \Sigma$ and ρ is ordinary at ℓ, i.e.,

$$\rho|D_\ell \cong \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}.$$

[[This definition could be completely wrong. It was definitely not presented clearly in class.]]

There is an exercise associated with this. It is to justify *a priori* the definition of δ. Suppose, for example, that $\rho_{f,\lambda}$ satisfies $(*)$, then we want to show that $\ell^2 \nmid N(f)$.

Theorem 17.3.3. *Every $\tilde{\rho}$ of class Σ comes from $S_2(\Gamma_0(N_\Sigma))$.*

Define an approximation $\mathbf{T}$ to the Hecke algebra by letting

$$\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots] \subset \operatorname{End}(S_2(\Gamma_0(N_\Sigma)))$$

where we adjoin only those T_n for which $(n, \ell N_\Sigma) = 1$. For some reason there exists a map

$$\mathbf{T} \to \mathbf{F}_{\ell^\nu} : \quad T_r \mapsto \operatorname{Tr}\rho(\operatorname{Frob}_r).$$

Why should such a map exist? The point is that we know by the theorem that ρ comes by reduction from a $\rho_{f,\lambda}$ with $f \in S_2(\Gamma_0(N_\Sigma))$.

But there is a wrinkle. [[I do not understand: He says: "Clearly $N(\rho)|N_\Sigma$. $k(\rho) = 2$ or $\ell+1$. If $k(\rho) = \ell+1$ then $\delta = 1$ so $\ell | N_\Sigma$. Have to trickily slip over to weight 2 in order to get f." This does not make any sense to me.]]

17.4 Wiles' Hecke algebra

Composing the map $\mathbf{T} \to \mathcal{O}_{E_f}$ with reduction mod λ from $\mathcal{O}_{E_f}$ to $\mathbf{F}_{\ell^\nu}$ we obtain a map $\mathbf{T} \to \mathbf{F}_{\ell^\nu}$. Let $\mathbf{m}$ be the kernel. Then $\mathbf{m}$ is a maximal ideal of $\mathbf{T}$. Wiles' Hecke algebra is the completion $\mathbf{T}_\mathbf{m}$ of $\mathbf{T}$ at $\mathbf{m}$. [[For some reason]] there exists

$$\tilde{\rho} : G = \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \operatorname{GL}(2, \mathbf{T}_\mathbf{m})$$

such that $\mathrm{Tr}(\tilde{\rho}(\mathrm{Frob}_r)) = T_r$. What makes this useful is that $\tilde{\rho}$ is *universal* for lifts of type Σ. This means that given any lift τ of type Σ there exists a map

$$\varphi : \mathrm{GL}(2, \mathbf{T}_{\mathbf{m}}) \to \mathrm{GL}(2, A)$$

such that $\tau = \varphi\tilde{\rho}$.

Another key idea that the approximation $\mathbf{T}$ obtained by just adjoining those T_n with $(n, \ell N_\Sigma) = 1$ is, after completing at certain primes, actually equal to the whole Hecke algebra.

Chapter 18

The Hecke Algebra T_Σ

18.1 The Hecke Algebra

Throughout this lecture $\ell \neq p$ and $\ell \geq 3$. We are studying the representation $\rho : G \to \mathrm{GL}(2, \mathbf{F}_{\ell^\nu})$. This is an irreducible representation, ℓ is odd, ρ is semistable, and $\det \rho = \chi$. To single out certain classes of liftings we let Σ be a finite set of primes. Let A be a complete local Noetherian ring with residue filed $\mathbf{F}_{\ell^\nu}$. We take liftings $\tilde{\rho} : G \to \mathrm{GL}(2, A)$ such that $\tilde{\rho}$ reduces down to ρ, $\det \tilde{\rho} = \tilde{\chi}$, and $\tilde{\rho}$ is "like" ρ away from Σ. For example, if ρ is unramified at p we also want $\tilde{\rho}$ unramified at p, etc.

Assume $\tilde{\rho}$ is modular. We guess the serious divisibility condition that $N(f)|N_\Sigma$. Recall that

$$N_\Sigma = \prod_{\substack{p\neq\ell \\ p\in\Sigma}} p^2 \cdot \prod_{\substack{p\neq\ell \\ p\notin\Sigma}} p^{\mathrm{ord}_p N(\rho)} \cdot \ell^\delta.$$

To define δ consider two cases.

- *level 1 case.* Take $\delta = 1$ if $\ell \in \Sigma$ or if ρ is not finite at ℓ. Take $\delta = 0$ otherwise.
- *level 2 case.* This is the case when $\rho|I_\ell$ has order $\ell^2 - 1$. Take $\delta = 0$.

A priori nobody seems to know how to prove that $N(f)|N_\Sigma$ given only that $\tilde{\rho}$ is modular. In the end we will show that all modular $\tilde{\rho}$ in fact come from

$$S_2(\Gamma_0(N_\Sigma)).$$

This can be regarded as a proof that $N(\rho)|N_\Sigma$.

Last time we tried to get things going by defining the *anemic Hecke algebra*

$$\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(S_2(\Gamma_0(N_\Sigma)))$$

where we only adjoin those T_n for which $(n, \ell N_\Sigma) = 1$. By some level lowering theorem there exists an $f \in S_w(\Gamma_0(N_\Sigma))$ giving $\tilde{\rho}$ so we obtain a map $\mathbf{T} \to \mathbf{F}$. The map sends T_n to the reduction modulo λ of its eigenvalue on f. (λ is a prime of the ring of integers of E_f lying over ℓ.) [[... something about needing an f of the right level $= N(\rho)$.]]

Let $\mathbf{m} \subset \mathbf{T}$ be the kernel of the above defined map $\mathbf{T} \to \mathbf{F}$. Then $\mathbf{m}$ is a maximal ideal. Let $\mathbf{T_m}$ be the completion of $\mathbf{T}$ at $\mathbf{m}$ and note that

$$\mathbf{T_m} \hookrightarrow \mathbf{T} \otimes_{\mathbf{Z}} \mathbf{Z}_\ell.$$

We need to know that $\mathbf{T}_\mathbf{m}$ is Gorenstein. This is done by comparing $\mathbf{T}_\mathbf{m}$ to some full Hecke ring. Thus

$$\mathbf{T} \subset R = \mathbf{Z}[\ldots, T_n, \ldots] \subset \mathrm{End}(S_2(\Gamma_0(N_\Sigma)))$$

where the full Hecke ring R is obtained by adjoining the T_n for *all* integers n. We should think of R as

$$R = \mathbf{T}[T_\ell, \{U_P : p|N_\Sigma\}].$$

Note that T_ℓ may or may not be a U_ℓ depending on if $\ell|N_\Sigma$.

Lemma 18.1.1. *If $\ell \nmid N_\Sigma$ then $R = \mathbf{T}[U_p, \ldots]$. Thus if T_ℓ is not a U_ℓ then we do not need T_ℓ.*

Proof. This lemma is an interesting thing and the proof goes as follows. Oooh. Sorry, this is not true. Ummm.... hmm.

The ring $\mathbf{T}[U_p, \ldots]$ is clearly of finite index in R since: if q is a random prime number consider $R \otimes_\mathbf{Z} \mathbf{Q}_q$ compared to $\mathbf{T}[U_p, \ldots] \otimes_\mathbf{Z} \mathbf{Q}_q$. [[I do not know how to do this argument. The lemma as stated above probably isn't really true. The point is that the following lemma is the one we need and it is true.]] □

Let $\mathbf{T}[U_p, \ldots]$ be the ring obtained by adjoining to $\mathbf{T}$ just those U_p with $p|N_\Sigma$.

Lemma 18.1.2. *If $\ell \nmid N_\Sigma$ then the index $(R : \mathbf{T}[U_p, \ldots])$ is prime to ℓ. Note that we assume $\ell \geq 3$.*

Proof. We must show that the map $\mathbf{T}[U_p, \ldots] \to R/\ell R$ is surjective. [[I thought about it for a minute and did not see why this suffices. Am I being stupid?]] Let $A = \mathbf{F}_\ell[T_n : (n, \ell) = 1]$ be the image of $\mathbf{T}[U_p, \ldots]$ in $R/\ell R$ so we have a diagram

$$\begin{array}{ccc} \mathbf{T}[U_p, \ldots] & \longrightarrow & R/\ell R \\ & \searrow & \cup \\ & & A \end{array}$$

We must show that $A = R/\ell R$. There is a beautiful duality

$$R/\ell R \times S_2(\Gamma_0(N_\Sigma); \mathbf{F}_\ell) \to \mathbf{F}_\ell \quad \text{(perfect pairing)}.$$

Thus $A^\perp = 0$ iff $A = R/\ell R$.

Suppose $0 \neq f \in A^\perp$, then $a_n(f) = 0$ for all n such that $(n, \ell) = 1$. Thus $f = \sum a_{n\ell} q^{n\ell}$. Let $\theta = q\frac{d}{dq}$ be the theta operator. Since the characteristic is ℓ, $\theta(f) = 0$. On the other hand $w(f) = 2$ and since $\ell \geq 3$, $\ell \nmid 2$. Thus $w(\theta f) = w(f) + \ell + 1 = 2 + \ell + 1$ which is a contradiction since $\theta f = 0$ and $w(0) = 0 \neq 3 + \ell$. [[The weight is an integer *not* a number mod ℓ, right?]] □

Example 18.1.3. The lemma only applies if $\ell \geq 3$. Suppose $\ell = 2$ and consider $S_2(\Gamma_0(23))$. Then

$$\mathbf{T}[U_p, \ldots] = \mathbf{Z}[\sqrt{5}] \subset R = \mathbf{Z}[\frac{1+\sqrt{5}}{2}]$$

so $(R : \mathbf{T}[U_p, \ldots])$ is not prime to 2.

Remark 18.1.4. If $N = N_\Sigma$ then

$$\operatorname{rank}_{\mathbf{Z}} \mathbf{T} = \sum_{M|N} S_2(\Gamma_0(M))^{\text{new}}$$

and

$$\operatorname{rank}_{\mathbf{Z}} R = \dim S_2(\Gamma_0(N)).$$

There is an injection

$$\bigoplus_{M|N} S_2(\Gamma_0(M))^{\text{new}} \hookrightarrow S_2(\Gamma_0(N))$$

but $\oplus S_2(\Gamma_0(M))^{\text{new}}$ is typically much smaller than $S_2(\Gamma_0(N))$.

18.2 The maximal ideal in R

The plan is to find a special maximal ideal $\mathbf{m}_R$

$$\begin{array}{ccc} \mathbf{m}_R & \subset & R \\ \cup & & \cup \\ \mathbf{m} & \subset & \mathbf{T} \end{array}.$$

Once we finally found the correct $\mathbf{m}$ and $\mathbf{m}_R$ then $\mathbf{T}_{\mathbf{m}} \xrightarrow{\sim} R_{\mathbf{m}_R}$ will be an isomorphism.

In finding $\mathbf{m}_R$ we will not invoke some abstract going up theorem but we will "produce" $\mathbf{m}_R$ by some other process.

The ideal $\mathbf{m}$ was defined by a newform f level $M|N_\Sigma$. The coefficients of f lie in

$$\mathcal{O}_\lambda = (\mathcal{O}_{E_f})_\lambda$$

where E_f is the coefficient ring of f and λ is a prime lying over ℓ. Thus $\mathcal{O}_\lambda$ is an ℓ-adic integer ring. Composing the residue class map $\mathcal{O}_\lambda \to \overline{\mathbf{F}}$ with the eigenvalue map $\mathbf{T} \to \mathcal{O}_\lambda$ we obtain the map $\mathbf{T} \to \overline{\mathbf{F}}$.

In order to obtain the correct $\mathbf{m}_R$ we will make a sequence of changes to f to make some good newform. [[Is the motivation for all this the fact that the lemma will not apply if $\ell | N_\Sigma$?]]

18.2.1 Strip away certain Euler factors

Write $f = \sum a_n q^n$. Replace f by

$$h = \sum_{\text{certain } n} a_n q^n$$

where the sum is over those n which are prime to each $p \in \Sigma$. What does this mean? If we think about the L-function $L(f,s) = \prod_p L_p(f,s)$, then h has L-function

$$L(h,s) = \prod_{p \notin \Sigma} L_p(f,s).$$

Furthermore making this change does not take us out of $S_2(\Gamma_0(N_\Sigma))$, i.e., $h \in S_2(\Gamma_0(N_\Sigma))$. [[He explained why but my notes are very incomplete. They say: Why. Because $h = (f \otimes \varepsilon) \otimes \varepsilon$, ($\varepsilon$ Dirichlet character ramified at primes in N_Σ). Get a form of level $\operatorname{lcm}(\prod_{p \in \Sigma} p^2, N(f)) | N_\Sigma$. Can strip and stay in space since N_Σ has correct squares built into it.]]

Remark 18.2.1. Suppose that $p||N_\Sigma$. Then $p \notin \Sigma$ and $p||N(\rho)$. The level of f is

$$N(f) = \begin{cases} N(\rho), & \text{if } k(\rho) = 2 \\ N(\rho)\ell, & \text{if } k(\rho) = \ell + 1 \end{cases}.$$

If $p||N(\rho)$ then $f|T_p = a_p(f)f$. Thus h is already an eigenform for T_p unless $p \in \Sigma$ in which the eigenvalue is 0.

[[I do not understand this remark. Why would the eigenvalue being 0 mean that T_p is not an eigenform? Furthermore, what is the point of this remark in the wider context of transforming f into a good newform.]]

18.2.2 Make into an eigenform for U_ℓ

We perform this operation to f to obtain a form g then apply the above operation to get the ultimate h having the desired properties. Do this if $\ell|N_\Sigma$ but $\ell \nmid N(f)$. This happens precisely if $\ell \in \Sigma$ and ρ is good and ordinary at ℓ. Then $f|U_\ell$ is just some random junk. Consider $g = f + *f(q^\ell)$ where $*$ is some coefficient. We see that if $*$ is chosen correctly then $g|U_\ell = Cg$ for some constant C. There are 2 possible choices for $*$ which lead to 2 choices for C. Let $a_\ell = a_\ell(f)$, then C can be either root of

$$X^2 + a_\ell X + \ell = 0.$$

This equation has exactly one unit root in $\mathcal{O}_\lambda$. The reason is because we are in the ordinary situation so a_ℓ is a unit. But ℓ is not a unit. The sum of the roots is a unit but the product is not. [[Even this is not clear to me right now. Definitely check this later.]] Make the choice of $*$ so that real root is C. Then we obtain a g such that

$$g|U_\ell = (\text{unit}) \cdot g.$$

Next apply the above procedure to strip g and end up with an h such that

- $h|U_\ell = (\text{unit}) \cdot h$,
- $h|U_p = 0$ for $p \in \Sigma$ $(p \neq \ell)$, and
- $h|T_p = a_p(f) \cdot h$ for $p \notin \Sigma$.

Now take the form h. It gives a map $R \to \mathcal{O}_\lambda$ which extends $\mathbf{T} \to \mathcal{O}_\lambda$. Let $\mathbf{m}_R$ be the kernel of the map $R \to \overline{\mathbf{F}}$ obtained by composing $R \to \mathcal{O}_\lambda$ with $\mathcal{O}_\lambda \to \overline{\mathbf{F}}$. A lot of further analysis shows that $\mathbf{T_m} \to R_{\mathbf{m}_r}$ is an isomorphism. We end up having to show separately that the map is injective and surjective.

[[In this whole lecture $p \neq \ell$.]]

[[Wiles' notation: His $\mathbf{T}_{\mathbf{m}_R}$ is my $\mathbf{R}$ and his $\mathbf{T}'$ is my $\mathbf{T_m}$.

18.3 The Galois Representation

We started with a representation ρ, chose a finite set of primes Σ and then made the completed Hecke algebra $\mathbf{T_m}$. Our goal is to construct the universal deformation of ρ of type Σ. The universal deformation is a representation

$$\tilde{\rho} : G \to \mathrm{GL}(2, \mathbf{T_m})$$

such that for all primes p with $p \nmid \ell N_\Sigma$, $\tilde{\rho}(\mathrm{Frob}_p)$ has trace $T_p \in \mathbf{T_m}$ and determinent p.

We now proceed with the construction of $\tilde{\rho}$. Let

$$\mathbf{T} = \mathbf{Z}[\ldots, T_n, \ldots], \quad (n, \ell N_\Sigma) = 1$$

be the anemic Hecke algebra. Then $\mathbf{T} \otimes \mathbf{Q}$ decomposes as a product of fields

$$\mathbf{T} \otimes \mathbf{Q} = \prod_f E_f$$

where the product is over a set of representatives for the Galois conjugacy classes of newforms of weight 2, trivial character, and level dividing N_Σ. Since $\mathbf{T}$ is integral (it is for example a finite rank $\mathbf{Z}$-module), $\mathbf{T} \hookrightarrow \prod_f \mathcal{O}_f$. Since $\mathbf{Z}_\ell$ is a flat $\mathbf{Z}$-module,

$$\mathbf{T} \otimes \mathbf{Z}_\ell \hookrightarrow \prod_f \mathcal{O}_f \otimes \mathbf{Z}_\ell = \prod_{f,\lambda} \mathcal{O}_{f,\lambda}$$

where the product is over a set of representatives f and all $\lambda | \ell$.

$\mathbf{T_m}$ is a direct factor of $\mathbf{T} \otimes \mathbf{Z}_\ell$. [[This is definitely not the assertion that $\mathbf{T_m}$ is an $\mathcal{O}_{f,\lambda}$. What exactly is it the assertion of really?]]

We can restrict the product to a certain finite set S and still obtain an injection

$$\mathbf{T_m} \hookrightarrow \prod_{(f,\lambda)\in S} \mathcal{O}_{f,\lambda}.$$

The finite set S consists of those (f, λ) such that the prime λ of $\mathcal{O}_f$ pulls back to $\mathbf{m}$ under the map $\mathbf{T} \to \mathcal{O}_f$ obtained by composing $\mathbf{T} \to \prod_f \mathcal{O}_f$ with the projection onto $\mathcal{O}_f$. [[Why is this enough so that $\mathbf{T_m}$ still injects in?]] Restricting to a finite product is needed so that

$$[\prod_{(f,\lambda)\in S} \mathcal{O}_{f,\lambda} : \mathbf{T_m}] < \infty.$$

Given f and λ there exists a representation

$$\rho_{f,\lambda} : G \to \mathrm{GL}(2, \mathcal{O}_{f,\lambda}).$$

It is such that $\mathrm{Tr}\, \rho_{f,\lambda}(\mathrm{Frob}_p) = a_p$ is the image of T_p under the inclusion

$$\mathbf{T} \otimes \mathbf{Z}_\ell \hookrightarrow \prod_{(f,\lambda)\in S} \mathcal{O}_{f,\lambda}.$$

Put some of these $\rho_{f,\lambda}$ together to create a new representation

$$\rho' = \prod_{(f,\lambda)\in S} \rho_{f,\lambda} : G \to \mathrm{GL}(2, \prod \mathcal{O}_{f,\lambda}) \subset \mathrm{GL}(2, \mathbf{T_m} \otimes \mathbf{Q}).$$

The sought after universal deformation $\tilde{\rho}$ is a map making the following diagram commute

$$\begin{array}{ccc} G & \xrightarrow{\rho'} & \mathrm{GL}(2, \mathbf{T_m} \otimes \mathbf{Q}) \\ & \tilde{\rho} \searrow & \cup \\ & & \mathrm{GL}(2, \mathbf{T_m}) \end{array}$$

Theorem 18.3.1. *ρ' is equivalent to a representation taking values in* $\mathrm{GL}(2, \mathbf{T_m})$.

One way to [[try to]] prove this theorem is by invoking a general theorem of Carayol. [[and then what? does this way work? why is it not a good way?]] But the right way to prove the theorem is Wiles' way.

18.3.1 The structure of $\mathbf{T_m}$

Just as an aside let us review the structure of $\mathbf{T_m}$.

- $\mathbf{T_m}$ is local.
- $\mathbf{T_m}$ is not necessarily a discrete valuation ring.
- $\mathbf{T_m} \otimes \mathbf{Q}$ is a product of finite extensions of $\mathbf{Q}_\ell$.
- $\mathbf{T_m}$ is not necessarily a product of rings $\mathcal{O}_{f,\lambda}$.
- $\mathbf{T_m}$ need not be integral. [[or can I say, "IS not integral." ?]]

18.3.2 The philosophy in this picture

Choose c to be a complex conjugation in $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Since ℓ is odd $\det(c) = -1$ is a very strong condition which rigidifies the situation.

18.3.3 Massage ρ

Choose two 1-dimensional subspaces so that

$$\rho(c) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

with respect to any basis consisting of one vector from each subspace. For any $\sigma \in G$ write

$$\rho(\sigma) = \begin{pmatrix} a_\sigma & b_\sigma \\ c_\sigma & d_\sigma \end{pmatrix}.$$

Then $a_\sigma d_\sigma$ and $b_\sigma c_\sigma$ are somehow intrinsically defined. This is because

$$\rho(\sigma c) = \begin{pmatrix} -a_\sigma & ? \\ ? & d_\sigma \end{pmatrix}$$

so

$$a_\sigma = \frac{\mathrm{Tr}(\rho(\sigma)) - \mathrm{Tr}(\rho(\sigma c))}{2}$$

and

$$d_\sigma = \frac{\mathrm{Tr}(\rho(\sigma)) + \mathrm{Tr}(\rho(\sigma c))}{2}.$$

Since we know the determinent it follows that $b_\sigma c_\sigma$ is also intrinsically known. [[The point is that we know certain things about these matrices in terms of their traces and determinants.]]

Proposition 18.3.2. *There exists $g \in G$ such that $b_g c_g \neq 0$.*

Proof. Since ρ is irreducible there exists σ_1 such that $b_{\sigma_1} \neq 0$ and there exists σ_2 such that $c_{\sigma_2} \neq 0$. If $b_{\sigma_2} \neq 0$ or $c_{\sigma_1} \neq 0$ then we are done. So the only problem case is when $b_{\sigma_1} = 0$ and $c_{\sigma_2} = 0$. Easy linear algebra shows that in this situation $g = \sigma_1\sigma_2$ has the required property. □

Now rigidify by choosing a basis so that $b_g = 1$. Doing this does not fix a basis because they are many ways to choose such a basis.

18.3.4 Massage ρ'

Choose a basis of $(\mathbf{T_m} \otimes \mathbf{Q})^2$ so that

$$\rho'(c) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

For any $\sigma \in G$ write

$$\rho'(\sigma) = \begin{pmatrix} a_\sigma & b_\sigma \\ c_\sigma & d_\sigma \end{pmatrix}$$

Using an argument as above shows that $a_\sigma, d_\sigma \in \mathbf{T_m}$ since the traces live in $\mathbf{T_m}$. Furthermore $b_\sigma c_\sigma \in \mathbf{T_m}$ since the determinent is in $\mathbf{T_m}$.

The *key observation* is that $b_\sigma c_\sigma$ reduces mod $\mathbf{m}$ to give the previous $b_\sigma c_\sigma \in \mathbf{F}$ corresponding to $\rho(\sigma)$. This is because the determinants and traces of ρ' are lifts of the ones from ρ. Since $\mathbf{m}$ is the maximal ideal of a local ring and $b_g c_g$ reduces mod $\mathbf{m}$ to something nonzero it follows that $b_g c_g$ is a unit in $\mathbf{T_m}$.

Choose a basis so that

$$\rho'(c) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

and also so that

$$\rho'(g) = \begin{pmatrix} a_g & 1 \\ u & d_g \end{pmatrix} \in \mathrm{GL}(2, \mathbf{T_m}).$$

Here u is a unit in $\mathbf{T_m}$.

Proposition 18.3.3. *Write*

$$\rho'(\sigma) = \begin{pmatrix} a_\sigma & b_\sigma \\ c_\sigma & d_\sigma \end{pmatrix}$$

with respect to the basis chosen above. Then $a_\sigma, b_\sigma, c_\sigma, d_\sigma \in \mathbf{T_m}$.

Proof. We already know that $a_\sigma, d_\sigma \in \mathbf{T_m}$. The question is how to show that $b_\sigma, c_\sigma \in \mathbf{T_m}$. Since

$$\rho'(\sigma g) = \begin{pmatrix} a_\sigma a_g + b_\sigma u & ? \\ ? & c_\sigma + d_\sigma d_g \end{pmatrix}$$

we see that $a_\sigma a_g + b_\sigma u \in \mathbf{T_m}$. Since $a_\sigma a_g \in \mathbf{T_m}$ it follows that $b_\sigma u \in \mathbf{T_m}$. Since u is a unit in $\mathbf{T_m}$ this implies $b_\sigma \in \mathbf{T_m}$. Similarly $c_\sigma + d_\sigma d_g \in \mathbf{T_m}$ so $c_\sigma \in \mathbf{T_m}$. $\square$

As you now see, in this situation we can prove Carayol's theorem with just some matrix computations. [[This is basically a field lowering representation theorem. The thing that makes it easy is that there exists something (namely c) with distinct eigenvalues which is rational over the residue field. Schur's paper, models over smaller fields. "Schur's method".]]

18.3.5 Representations from modular forms mod ℓ

If you remember back in the 70's people would take an $f \in S_2(\Gamma_0(N); \mathbf{F})$ which is an eigenform for almost all the Hecke operators

$$T_p f = c_p f \quad \text{for almost all } p, \text{ and } c_p \in \mathbf{F}.$$

The question is then: Can you find

$$\rho : G \to \mathrm{GL}(2, \mathbf{F})$$

such that

$$\mathrm{Tr}(\rho(\mathrm{Frob}_p)) = c_p \text{ and } \det(\rho(\mathrm{Frob}_p)) = p$$

for all but finitely many p? The answer is yes. The idea is to find ρ by taking $\rho_{f,\lambda}$ [[which was constructed by Shimura?]] for some f and reducing mod λ. The only special thing that we need is a lemma saying that the eigenvalues in characteristic ℓ lift to eigenvalues in characteristic 0.

18.3.6 Representations from modular forms mod ℓ^n

Serre and Deligne asked: "What happens mod ℓ^n?"

More precisely, let R be a local finite Artin ring such that $\ell^n R = 0$ for some n. Take $f \in S_2(\Gamma_0(N); R)$ satisfying the hypothesis

$$\{r \in R : rf = 0\} = \{0\}.$$

This is done to insure that certain eigenvalues are unique. Assume that for almost all p one has $T_p f = c_p f$ with $c_p \in R$. The problem is to find $\rho : G \to \mathrm{GL}(2, R)$ such that

$$\mathrm{Tr}(\rho(\mathrm{Frob}_p)) = c_p \text{ and } \det(\rho(\mathrm{Frob}_p)) = p$$

for almost all p.

The big stumbling block is that ρ need not be the reduction of some $\rho_{g,\lambda}$ for any g, λ. [[I couldn't understand why – I wrote "can mix up f's from characteristic 0 so can not get one which reduces correctly."]]

Let $\mathbf{T} = \mathbf{Z}[\ldots, T_p, \ldots]$ where we only adjoin those T_p for which f is an eigenvector [[I made this last part up, but it seems very reasonable]]. Then f obviously gives a rise to a map

$$\mathbf{T} \to R : T \mapsto \text{ eigenvalue of } T \text{ on } f\ .$$

The strange hypothesis on f insures that the eigenvalue is unique. Indeed, suppose $Tf = af$ and $Tf = bf$, then $0 = Tf - Tf = af - bf = (a - b)f$ so by the hypothesis $a - b = 0$ so $a = b$.

Since the pullback of the maximal ideal of R is a maximal ideal of $\mathbf{T}$ we get a map $\mathbf{T}_\mathbf{m} \to R$ for some $\mathbf{m}$. [[I do not understand why we suddenly get this map and I do not know why the pullback of the maximal ideal is maximal.]]

Now the problem is solved. Take $\rho' : G \to \mathrm{GL}(2, \mathbf{T}_\mathbf{m})$ with the sought after trace and determinent properties. Then let ρ be the map obtained by composing with the map $\mathrm{GL}(2, \mathbf{T}_\mathbf{m}) \to \mathrm{GL}(2, R)$.

18.4 ρ' is of type Σ

Let ρ be modular irreducible and semistable mod ℓ representation with $\ell > 2$. Let Σ be a finite set of primes. Then $N(\rho) | N_\Sigma$. We constructed the anemic Hecke algebra $\mathbf{T}$ which contains a certain maximal ideal $\mathbf{m}$. We then consider the completion $\mathbf{T}_\mathbf{m}$ of $\mathbf{T}$ at $\mathbf{m}$. Next we constructed

$$\rho' : G \to \mathrm{GL}(2, \mathbf{T}_\mathbf{m})$$

lifting ρ. Thus the diagram

$$\begin{array}{ccc} G & \xrightarrow{\rho'} & GL(2, \mathbf{T}_\mathbf{m}) \\ & \rho \searrow & \downarrow \\ & & GL(2, \mathbf{F}) \end{array}$$

commutes.

Some defining properties of ρ' are

- $\det \rho' = \tilde{\chi}$.
- $\operatorname{Tr}\rho'(\mathrm{Frob}_r) = T_r$. Since topologically $\mathbf{T_m}$ is generated by the Frob_r this is a tight condition.
- ρ' is a lift of type Σ.

To say ρ' is a lift of type Σ entails that ρ' is unramified outside primes $p|N_\Sigma$. This is true because ρ' is constructed from various $\rho_{f,\lambda}$ with $N(f)|N_\Sigma$. If $p \neq \ell$ and $p|N(\rho)$ then $p||N_\Sigma$. Recall that

$$\rho|D_p \sim \begin{pmatrix} \alpha & * \\ 0 & \beta \end{pmatrix}$$

where $\alpha = \beta\chi$ and α and β are unramified. For ρ' to be a lift of type Σ we require that

$$\rho'|D_p \sim \begin{pmatrix} \tilde{\alpha} & * \\ 0 & \tilde{\beta} \end{pmatrix}$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are unramified lifts and $\tilde{\alpha} = \tilde{\beta}\chi$. [[I find it mighty odd that α is χ times an unramified character and yet α is not ramified! How can that be? Restricted to inertia β and α would be trivial but χ would not be.]] Is this true of ρ'? Yes since by a theorem of Langlands the factors

$$\rho_{f,\lambda}|D_p \sim \begin{pmatrix} \tilde{\alpha} & * \\ 0 & \tilde{\beta} \end{pmatrix}.$$

[[Ribet said more about this but it does not form a cohesive whole in my mind. Here is what I have got. Since $\rho_{f,\lambda}$ obviously ramified at p, $p||N(f)|N_\Sigma$. $\rho_{f,\lambda}|D_p$ is like an elliptic curve with bad multiplicative reduction at p. That $\rho_{f,\lambda}|D_p \sim \left(\begin{smallmatrix} \tilde{\alpha} & * \\ 0 & \tilde{\beta} \end{smallmatrix}\right)$ really comes down to Deligne-Rapaport. If write $f = \sum a_n q^n$, then $a_p \neq 0$ and $\tilde{\beta}(\mathrm{Frob}_p) = a_p$, $\tilde{\alpha}(\mathrm{Frob}_p) = pa_p$. Thus $a_p^2 = 1$ since $\tilde{\alpha}\tilde{\beta} = \chi$. Thus $a_p = \pm 1$ and $a_p \mod \lambda = \beta(\mathrm{Frob}_p) = \pm 1$ independent of (f, λ). So we have these numbers $a_p = a_p(f) = \pm 1$, independent of λ.]]

18.5 Isomorphism between $\mathbf{T_m}$ and $R_{\mathbf{m}_R}$

Let $\mathbf{T} \subset R = \mathbf{Z}[\ldots, T_n, \ldots]$ be the anemic Hecke algebra with maximal ideal $\mathbf{m}$. The difference between $\mathbf{T}$ and R is that R contains all the Hecke operators whereas $\mathbf{T}$ only contains the T_p with $p \not| \ell N_\Sigma$. Wiles proved that the map $\mathbf{T_m} \to R_{\mathbf{m}_R}$ is an isomorphism. Which Hecke operators are going to hit the missing T_p? If we do the analysis in $R_{\mathbf{m}_R}$ we see that [[I think for $p \neq \ell$!]]

$$T_p = \begin{cases} \pm 1, & \text{for } p|N_\Sigma,\ p \notin \Sigma \\ 0, & \text{for } p \in \Sigma \end{cases}.$$

This takes care of everything except T_ℓ. In proving the surjectivity of $\mathbf{T_m} \to R_{\mathbf{m}_R}$ we are quite happy to know that $T_p = \pm 1$ or 0. Proving this is a bit unpalatable. It is described in [4].

Consider the commuting diagram

$$\begin{array}{ccc} \mathbf{T_m} & \hookrightarrow & \prod \mathcal{O}_{f,\lambda} \\ & \searrow & \uparrow \\ & & R_{\mathbf{m}_R} \end{array}.$$

The map $R_{\mathbf{m}_R} \to \prod \mathcal{O}_{f,\lambda}$ is constructed by massaging f by stripping away certain Euler factors so as to obtain an eigenvector for all the Hecke operators. This diagram forces $\mathbf{T_m} \to R_{\mathbf{m}_R}$ to be injective.

[[Ogus: Is it clearly surjective on the residue field? Ribet: Yes. Ogus: OK, then we just need to prove it is ètale.]]

From the theory of the θ operator we already know two-thirds of the times that $\mathbf{T}$ contains T_ℓ.

Suppose $\ell | N_\Sigma$. This entails that we are in the ordinary case, ρ is not finite at ℓ, or $\ell \in \Sigma$. We did not prove in this situation that $T_\ell \in \mathbf{Z}[\ldots, T_n, \ldots : (n,\ell)=1]$.

Using generators and relations and brute force one shows that $R_{\mathbf{m}_R} \to \prod \mathcal{O}_{f,\lambda}$ is an injection. Then we can compare everything in $\prod \mathcal{O}_{f,\lambda}$. Now

$$\rho_{f,\lambda}|D_\ell \sim \begin{pmatrix} \tilde\alpha & * \\ 0 & \tilde\beta \end{pmatrix}$$

and

$$\tilde\beta(\mathrm{Frob}_\ell) = T_\ell \in \prod \mathcal{O}_{f,\lambda}.$$

Using arguments like last time one shows that $\tilde\beta(\mathrm{Frob}_\ell)$ can be expressed in terms of the traces of various operators. This proves surjectivity in this case.

Ultimately we have $\mathbf{T_m} \cong R_{\mathbf{m}_R}$. The virtue of $\mathbf{T_m}$ is that it is generated by traces. The virtue of $R_{\mathbf{m}_R}$ is that it is Gorenstein. We have seen this if $\ell \not| N_\Sigma$. In fact it is Gorenstein even if $\ell || N_\Sigma$. [[Ribet: As I stand here today I do not know how to prove this last assertion in exactly one case. Ogus: You mean there is another gap in Wiles' proof. Ribet: No, it is just something I need to work out.]] When $\ell || N_\Sigma$ there are 2 cases. Either ρ is not finite at ℓ or it is. The case when ρ is not finite at ℓ was taken care of in [[cite a paper by Mazur-Ribet, Asterisque, reference?]]. A proof that $R_{\mathbf{m}_R}$ is Gorenstein when ρ is finite at ℓ ($\ell \in \Sigma$) is not in the literature.

Now forget $R_{\mathbf{m}_R}$ and just think of $\mathbf{T_m}$ in both ways: trace generated and Gorenstein.

18.6 Deformations

Fix an absolutely irreducible modular mod ℓ representation ρ and a finite set of primes Σ. Consider the category $\mathcal{C}$ of complete local Noetherian $W(\mathbf{F})$-algebras A (with $A/\mathbf{m} = \mathbf{F}$). Here $\mathbf{F} = \mathbf{F}_{\ell^\nu}$ and $W(\mathbf{F})$ is the ring of Witt vectors, i.e., the ring of integers of an unramified extension of $\mathbf{Q}_\ell$ of degree ν.

Define a functor $\mathcal{F} : \mathcal{C} \to \mathbf{Set}$ by sending A in $\mathcal{C}$ to the set of equivalence classes of lifts

$$\tilde\rho : G \to \mathrm{GL}(2, A)$$

of ρ of type Σ. The equivalence relation is that $\tilde\rho_1 \sim \tilde\rho_2$ iff there exists $M \in \mathrm{GL}(2,A)$ with $M \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{\mathbf{m}}$ such that $\tilde\rho_1 = M^{-1}\tilde\rho_2 M$.

Mazur proved [15] that $\mathcal{F}$ is representable.

Theorem 18.6.1. *There exists a lift*

$$\rho_{univ} : G \to \mathrm{GL}(2, R_\Sigma)$$

of type Σ such that given any lift $\tilde\rho : G \to \mathrm{GL}(2,A)$ there exists a unique homomorphism $\varphi : R_\Sigma \to A$ such that $\varphi \circ \rho_{univ} \sim \tilde\rho$ in the sense of the above equivalence relation.

Lenstra figured out how to concretely construct R_Σ.

Back in the student days of Ribet and Ogus, Schlessinger wrote a widely quoted thesis which gives conditions under which a certain class of functors can be representable. Mazur checks these conditions in his paper.

[[Buzzard: What happened to Schlessinger anyways? Ribet: He ended up at University of North Carolina, Chapel Hill.]]

We have constructed $\tilde{\rho} = \rho' : G \to \mathrm{GL}(2, \mathbf{T_m})$. By the theorem there exists a unique morphism $\varphi : R_\Sigma \to \mathbf{T_m}$ such that $\rho' = \varphi \circ \rho_{\mathrm{univ}}$.

Theorem 18.6.2. *φ is an isomorphism thus ρ' is the universal deformation and $\mathbf{T_m}$ is the universal deformation ring.*

This will imply that any lift of type Σ is modular.

The morphism φ is surjective since

$$T_p = \mathrm{Tr}\,\tilde{\rho}(\mathrm{Frob}_p) = \mathrm{Tr}\,\varphi \circ \rho_{\mathrm{univ}}(\mathrm{Frob}_p) = \varphi(\mathrm{Tr}(\rho_{\mathrm{univ}}(\mathrm{Frob}_p))).$$

We have two very abstractly defined local Noetherian rings. How would you prove they are isomorphic? Most people would be terrified by this question. Wiles dealt with it.

18.7 Wiles Main Conjecture

> "We are like a train which is trying to reach Fermat's Last Theorem. Of course it has not made all of its scheduled stops. But it is on its way."

We have a representation $\rho : G \to \mathrm{GL}(2, \mathbf{F})$. Take $\mathbf{F} = \mathbf{F}_\ell$ for our applications today. Then the ring of Witt vectors is $W(\mathbf{F}) = \mathbf{Z}_\ell$. The Hecke algebra can be embedded as

$$\mathbf{T_m} \subset \prod_{(g,\mu)\in\mathcal{A}} \mathcal{O}_{g,\mu}.$$

The Hecke algebra $\mathbf{T_m}$ has the following properties.

- The index of $\mathbf{T_m}$ in $\prod \mathcal{O}_{g,\mu}$ is finite.
- Gorenstein as a $\mathbf{Z}_\ell$-module, i.e., there exists an isomorphism $\mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{T_m}, \mathbf{Z}_\ell) \cong \mathbf{T_m}$.
- $\mathbf{T_m}$ is generated by the T_r with r prime and $(r, \ell N_\Sigma) = 1$.

We have constructed a representation

$$\rho' : G \to \mathrm{GL}(2, \mathbf{T_m}).$$

Composing appropriately with the map $\mathbf{T_m} \hookrightarrow \prod \mathcal{O}_{g,\mu}$ gives a map

$$G \to \prod_{(g,\mu)} \mathrm{GL}(2, \mathcal{O}_{g,\mu}).$$

This is the product of representations $\prod \rho_{g,\mu}$. The triangle is

$$\begin{array}{ccc} G & \xrightarrow{\rho'} & \mathrm{GL}(2, \mathbf{T_m}) \\ & \prod \rho_{g,\mu} \searrow & \downarrow \\ & & \prod \mathrm{GL}(2, \mathcal{O}_{g,\mu}) \end{array}$$

Moreover, ρ' is a deformation of ρ of type Σ so it lifts ρ and satisfies certain "nice as ρ" properties at primes $p \notin \Sigma$.

Let

$$\rho_{\mathrm{univ}} : G \to \mathrm{GL}(2, R_\Sigma)$$

be the universal deformation of ρ of type Σ. Lenstra gave a very concrete paper [12] constructing this ρ_{univ}. Before his paper there was only Schlesinger's thesis. By the definition of ρ_{univ} there exists a unique map $\varphi : R_\Sigma \to \mathbf{T_m} = \mathbf{T}_\Sigma$ such that $\varphi \circ \rho_{\mathrm{univ}} = \rho'$. By $\varphi \circ \rho_{\mathrm{univ}}$ we mean the composition of ρ_{univ} with the map $\mathrm{GL}(2, R_\Sigma) \to \mathrm{GL}(2, \mathbf{T_m})$ induces by φ. As noted last time it is easy to see that φ is surjective.

Theorem 18.7.1 (Wiles main 'conjecture'). *φ is an isomorphism (for each Σ).*

The theorem implies the following useful result.

Theorem 18.7.2. *Suppose E is a semistable elliptic curve over $\mathbf{Q}$ and that for some $\ell > 2$ the representation $\rho = \rho_{\ell,E}$ on $E[\ell]$ is irreducible and modular. Then E is modular.*

Proof. The representation

$$\tilde{\rho} = \rho_{\ell^\infty,E} : G \to \mathrm{GL}(2, \mathbf{Z}_\ell)$$

on the ℓ-power torsion $E[\ell^\infty] = \cup E[\ell^n]$ is a lift of

$$\rho : G \to \mathrm{Aut}(E[\ell]) = \mathrm{GL}(2, \mathbf{F}_\ell).$$

Furthermore, $\tilde{\rho}$ is a deformation of ρ of type Σ. Applying universality and using the theorem that $\mathbf{R}_\Sigma \cong \mathbf{T}_\Sigma$ we get a map

$$\mathbf{T}_\Sigma \to \mathbf{Z}_\ell : \quad T_r \mapsto a_r = a_r(E) = \mathrm{trace}\,\tilde{\rho}(\mathrm{Frob}_r).$$

The relevant diagram is

$$\begin{array}{ccc} R_\Sigma & & \\ \downarrow & \searrow & \\ \mathbf{T}_\Sigma & \longrightarrow & \mathbf{Z}_\ell \end{array}$$

where the map $R_\Sigma \to \mathbf{Z}_\ell$ is given by

$$\mathrm{trace}\,\rho_{\mathrm{univ}}(\mathrm{Frob}_r) \mapsto a_r.$$

Now the full Hecke algebra $\mathbf{Z}[\ldots, T_n, \ldots]$ embeds into the completion $\mathbf{T}_\Sigma$. Composing this with the map $\mathbf{T}_\Sigma \to \mathbf{Z}_\ell$ above we obtain a map

$$\alpha : \mathbf{Z}[\ldots, T_n, \ldots] \to \mathbf{Z}_\ell.$$

Because of the duality between the Hecke algebra and modular forms there exists a modular form $h \in S_2(\Gamma_0(N_\Sigma), \mathbf{Z}_\ell)$ corresponding to α. Since α is a homomorphism h is a normalized eigenform. Furthermore $a_r(h) = a_r(E) \in \mathbf{Z}$ for all primes $r \nmid \ell N_\Sigma$. Since almost all coefficients of h are integral it follows that h is integral. Because we know a lot about eigenforms we can massage h to an eigenform in $S_2(\Gamma_0(N_E), \mathbf{Z})$. $\square$

[[Some undigested comments follow.]]

- Once there is any connection between ρ and a modular form one can prove Taniyama-Shimura in as strong a form as desired. See the article *Number theory as Gadfly.*

- Take the abelian variety A_h attached to h. The λ-adic representation will have pieces with the same representations. Using Tate's conjecture we see that E is isogenous to A_h. Use at some points Carayol's theorem: If g is a form giving rising to the abelian variety A then the conductor of A is the same as the conductor of g.
- Tate proved that if two elliptic curves have isomorphic ρ_{ℓ^∞} for some ℓ then they are isogenous.

18.8 $\mathbf{T}_\Sigma$ is a complete intersection

Recall the construction of $\mathbf{T}_\Sigma = \mathbf{T}_\mathbf{m}$. Let $\mathbf{T}$ be the anemic Hecke algebra. Then

$$\mathbf{T} \otimes \mathbf{Z}_\ell \hookrightarrow \prod \mathcal{O}_{g,\mu}$$

where the product is over a complete set of representatives (for the action of Galois on eigenforms) g and primes μ lying over ℓ. We found a specific (f, λ) for $\Sigma = \emptyset$ such that $\overline{\rho}_{\lambda,f} = \rho$. The maximal ideal $\mathbf{m}$ was defined as follows. The form f induces a map $\mathbf{T} \to \mathcal{O}_{f,\lambda}$. Taking the quotient of $\mathcal{O}_{f,\lambda}$ by its maximal ideal we obtain a map $\mathbf{T} \to \overline{\mathbf{F}}_\ell$. Then $\mathbf{m}$ is the kernel of this map. The diagram is

$$\begin{array}{ccc} \mathbf{T} & & \\ \downarrow & \searrow & \\ \mathcal{O}_{f,\lambda} & \longrightarrow & \overline{\mathbf{F}}_\ell \end{array}$$

The map $\mathcal{O}_{f,\lambda} \to \overline{\mathbf{F}}_\ell$ is

$$a_r(f) \mapsto \operatorname{trace} \rho(\mathrm{Frob}_r).$$

To fix ideas we cheat and suppose $\mathcal{O}_{f,\lambda} = \mathbf{Z}_\ell$. [[In Wiles' optic this is OK since he can work this way then tensor everything at the end.]]

Now $\rho_{f,\lambda}$ is a distinguished deformation of ρ [["Distinguished" is not meant in a mathematical sense]]. The map f gives rise to a map $\mathbf{T}_\Sigma \to \mathcal{O} = \mathbf{Z}_\ell$ which we also denote by f

$$\begin{array}{ccc} R_\Sigma & \xrightarrow{\varphi} & \mathbf{T}_\Sigma \\ & \searrow & \downarrow f \\ & & \mathcal{O} = \mathbf{Z}_\ell \end{array}$$

18.9 The inequality $\#\mathcal{O}/\eta \leq \#\wp_T/\wp_T^2 \leq \wp_R/\wp_R^2$

Let

$$\rho : G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}(2, \mathbf{F}_\ell)$$

be irreducible and modular with $\ell > 2$. Let Σ be a finite set of primes. We assume there is a modular form f of weight 2 with coefficients in $\mathbf{Z}_\ell$ which gives rise ρ. Let

$$\rho_{f,\lambda} : G \to \mathrm{GL}(2, \mathbf{Z}_\ell)$$

be the representation coming from f, then $\rho_{f,\lambda}$ reduces to ρ modulo ℓ.

Let R_Σ be the universal deformation ring, so every deformation of ρ of type Σ factors through R_Σ in an appropriate sense. Let $\mathbf{T}_\Sigma$ be the Hecke ring associated to Σ. It is a $\mathbf{Z}_\ell$-algebra which is free of finite rank. Furthermore

$$\mathbf{T}_\Sigma \subset \prod_{(g,\mu)\in\mathcal{A}} \mathcal{O}_{g,\mu} = \mathcal{O}_{f,\mu} \times \prod_{(g,\mu)\in\mathcal{A}-\{(f,\mu)\}} \mathcal{O}_{g,\mu}$$

where $\mathcal{A}$ is as defined before. Define projections pr_1 and pr_2 onto the first and rest of the factors, respectively

$$\begin{aligned} pr_1 &: \prod_{(g,\mu)\in\mathcal{A}} \mathcal{O}_{g,\mu} \to \mathcal{O} = \mathcal{O}_{f,\lambda} \\ pr_2 &: \prod_{(g,\mu)\in\mathcal{A}} \mathcal{O}_{g,\mu} \to \prod_{(g,\mu)\neq(f,\lambda)} \mathcal{O}_{g,\mu} \end{aligned}$$

Let $\varphi : R_\Sigma \to \mathbf{T}_\Sigma$ be the map coming from the universal property of R_Σ. This map is surjective. The famous triangle which dominates all of the theory is

$$\begin{array}{ccc} R_\Sigma & \xrightarrow{\varphi}\!\!\!\!\rightarrow & \mathbf{T}_\Sigma \\ & & \downarrow pr_1 \\ & & \mathcal{O} = \mathbf{Z}_\ell \end{array}$$

18.9.1 The definitions of the ideals

We now define two ideals. View $\mathbf{T}_\Sigma$ as sitting in the product $\prod \mathcal{O}_{g,\mu}$.

1. The *congruence ideal* $\eta \subset \mathcal{O}$ is

$$\eta := \mathcal{O} \cap \mathbf{T}_\Sigma = \ker\Big(pr_2 : \mathbf{T}_\Sigma \to \prod_{(g,\mu)\neq(f,\lambda)} \mathcal{O}_{g,\mu}\Big)$$

2. The prime ideal $\wp_T \subset \mathbf{T}_\Sigma$ is

$$\wp_T = \ker\Big(pr_1 : \mathbf{T}_\Sigma \to \mathcal{O}\Big)$$

It is true that

$$\#\mathcal{O}/\eta \leq \#\wp_T/\wp_T^2.$$

The condition for equality is a theorem of Wiles.

Theorem 18.9.1. $\mathbf{T}_\Sigma$ *is a complete intersection iff* $\#\mathcal{O}/\eta = \#\wp_T/\wp_T^2$.

There is an analogous construction for

$$\psi = pr_1 \circ \varphi : R_\Sigma \to \mathcal{O}.$$

The diagram is

$$\begin{array}{ccc} R_\Sigma & \xrightarrow{\varphi} & \mathbf{T}_\Sigma \\ & \psi \searrow & \downarrow pr_1 \\ & & \mathcal{O} \end{array}.$$

Let $\wp_R$ be the kernel of ψ. From the commutativity of the above diagram we see that ψ maps $\wp_R \longrightarrow\!\!\!\!\rightarrow \wp_T$. Thus we have an induced map $\overline{\psi}$ on "tangent spaces"

$$\overline{\psi} : \wp_R/\wp_R^2 \longrightarrow\!\!\!\!\rightarrow \wp_T/\wp_T^2.$$

It follows that

$$\#\mathcal{O}/\eta \leq \#\wp_T/\wp_T^2 \leq \#\wp_R/\wp_R^2.$$

There is an analogous theorem.

Theorem 18.9.2. *The above inequalities are all equalities iff*

- $\varphi : R_\Sigma \to \mathbf{T}_\Sigma$ *is an isomorphism, and*
- $\mathbf{T}_\Sigma$ *is a complete intersection ring.*

18.9.2 Aside: Selmer Groups

Let M be the set of matrices in $M(2, \mathbf{Q}_\ell/\mathbf{Z}_\ell)$ which have trace 0. Then $\mathrm{GL}(2,\mathbf{Z}_\ell)$ operates on M by conjugation. Thus G acts on M via the representation $\rho' : G \to \mathrm{GL}(2,\mathbf{Z}_\ell)$. To $\wp_R/\wp_R^2$ there corresponds the *Selmer group* which is a subgroup of $H^1(\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}), M)$. The subgroup is

$$H^1_\Sigma(G, M) = \mathrm{Hom}_{\mathcal{O}}(\wp_R/\wp_R^2, \mathbf{Q}_\ell/\mathbf{Z}_\ell) \subset H^1(\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}), M).$$

[[Since Flach's thesis there has been a problem of trying to get an upper bound for the Selmer group. Wiles converted it into the above problem.]]

18.9.3 Outline of some proofs

We outline the key steps in the proof that $\#\wp_R/\wp_R^2 \le \#\mathcal{O}/\eta$.

Step 1: $\Sigma = \emptyset$

The key step is the minimal case when $\Sigma = \emptyset$. This is done in [35]. They claim to be proving the apparently weaker statement that $\mathbf{T}_\Sigma$ a complete intersection implies

$$\#\wp_T/\wp_T^2 = \#\mathcal{O}/\eta.$$

But in Wiles' paper [36] he obtains the inequality

$$\#\wp_R/\wp_R^2 \le \frac{(\#\wp_T/\wp_T^2)^2}{\#\mathcal{O}/\eta}.$$

Combining these two shows that

$$\#\wp_R/\wp_R^2 \le \#\mathcal{O}/\eta.$$

In an appendix to [35] Faltings proves directly that

$$\#\wp_R/\wp_R^2 \le \#\mathcal{O}/\eta.$$

[[At this point there were some remarks about why Wiles might have taken a circular route in his Annals paper. Ribet replied,

> "As Serre says, it is sometimes better to leave out any psychological behavior related to how people did something but instead just report on what they did."]]

Step 2: Passage from $\Sigma = \emptyset$ to σ general

The second step is the induction step in which we must understand what happens as Σ grows. Thus Σ is replaced by $\Sigma' = \Sigma \cup \{q\}$ where q is some prime not in Σ.

We will use the following notation. The object attached to Σ' will be denoted the same way as the object attached to Σ but with a $'$. Thus $(\wp_R/\wp_R^2)'$ denotes the Selmer group for Σ'.

The change in the Selmer group when Σ is replaced by Σ' is completely governed by some local cohomology group. There is a constant c_q such that

$$\#(\wp_R/\wp_R^2)' \le c_q \#(\wp_R/\wp_R^2) \le c_q \#\mathcal{O}/\eta.$$

So we just need to know that
$$\#\mathcal{O}/\eta' \geq c_q \#\mathcal{O}/\eta,$$
i.e., that η' is *small* as an ideal in $\mathcal{O}$. We need a formula for the ratio of the two orders.

Let $\mathbf{T}$ be the anemic ring of Hecke operators on $S_2(\Gamma_0(N_\Sigma))$ obtained by adjoining to $\mathbf{Z}$ all the Hecke operators $\mathbf{T}_n$ with n prime to ℓN_Σ. Let $\mathbf{T}'$ be the anemic Hecke ring of Hecke operators on $S_2(\Gamma_0(N_{\Sigma'}))$.

Since $N_\Sigma | N_{\Sigma'}$ there is an inclusion
$$S_2(\Gamma_0(N_\Sigma)) \hookrightarrow S_2(\Gamma_0(N_{\Sigma'})).$$
There is one subtlety, this injection is not equivariant for all of the Hecke operators. But this is no problem because $\mathbf{T}$ and $\mathbf{T}'$ are anemic. So the inclusion induces a restriction map $r : \mathbf{T}' \to \mathbf{T}$.

We now introduce a relative version of η which is an ideal $I \subset \mathbf{T}$. One way to think of I is as $\mathbf{T} \cap \mathbf{T}'$ where $\mathbf{T}$ and $\mathbf{T}'$ are both viewed as subrings of $\prod \mathcal{O}_{g,\mu}$
$$\begin{array}{ccc}
\mathbf{T}' & \xrightarrow{r} & \mathbf{T} \\
\cap & & \cap \\
\mathbf{T}'_{\mathbf{m}'} & & \mathbf{T}_{\mathbf{m}} \\
\cap & & \cap \\
\prod_{(g,\mu)} \mathcal{O}_{g,\mu} & \hookrightarrow & \prod_{\text{more } (g,\mu)} \mathcal{O}_{g,\mu}
\end{array}$$
The definition Lenstra would give is that
$$I := r(\mathrm{Ann}_{\mathbf{T}'}(\ker(r))).$$

The amazing formula is
$$\eta' = \eta \cdot f(I)$$
where $f : \mathbf{T} \to \mathcal{O}$ is the map induced by the modular form f. [[After introducing this definition Ogus was very curious about how deep it is, in particular, about whether its proof uses the Gorensteiness of $\mathbf{T}$. Ribet said, "somehow I do not think this formula can possibly be profound."]]

We pause with an aside to consider Wiles' original definition of η. By duality the map $f : \mathbf{T}_\Sigma \to \mathbf{Z}_\ell$ induces
$$f^\vee : \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{Z}_\ell, \mathbf{Z}_\ell) \to \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{T}_\Sigma, \mathbf{Z}_\ell) \cong \mathbf{T}_\Sigma.$$
Because $\mathbf{T}_\Sigma$ is Gorenstein there is an isomorphism $\mathrm{Hom}_{\mathbf{Z}_\ell}(\mathbf{T}_\Sigma, \mathbf{Z}_\ell) \cong \mathbf{T}_\Sigma$. Now $f^\vee(\mathrm{id}) \in \mathbf{T}_\Sigma$ so $f(f^\vee(\mathrm{id})) \in \mathcal{O} = \mathbf{Z}_\ell$. Wiles let $\eta = (f(f^\vee(\mathrm{id})))$ be the ideal generated by $f(f^\vee(\mathrm{id}))$.

To finish step 2 we must show that $\#\mathcal{O}/f(I) \geq c_q$, i.e., that "$I$ is small". [[I do not see how this actually finishes step 2, but it is reasonable that it should. How does this index relate to the index of $f(I)\eta$ in $\mathcal{O}$?]]

Let $J = J_0(N_\Sigma)$ and $J' = J_0(N_{\Sigma'})$. Since
$$S_2(\Gamma_0(N_\Sigma)) \hookrightarrow S_2(\Gamma_0(N_{\Sigma'}))$$
functoriality of the Jacobian induces a map $J \hookrightarrow J'$. By autoduality we also obtain an injection $J^\vee \hookrightarrow J'$ and $J \cap J^\vee$ is a finite subgroup of J'. [[I definitely do not understand why J is not just equal to $J^\perp$. Where does the other embedding $J^\perp \hookrightarrow J'$ come from?]]

It can be seen that $J \cap J^\vee = J[\delta]$ for some $\delta \in \mathbf{T}$. It turns out that
$$\mathrm{Ann}_{\mathbf{T}}(J \cap J^\vee) = \mathbf{T} \cap \delta\,\mathrm{End}(J) \supseteq \delta\mathbf{T}.$$

It is an observable fact that $f(\delta\mathbf{T})$ is an ideal of $\mathcal{O}$ of norm c_q. The *heart* of the whole matter is to see that the inclusion

$$\delta\mathbf{T} \subseteq \mathbf{T} \cap \delta\,\mathrm{End}(J)$$

is an equality after localization at $\mathbf{m}$. To do this we have to know that $\mathrm{Tate}_{\mathbf{m}}(J) \cong \mathbf{T}_{\mathbf{m}}^2$. This is equivalent to the Gorenstein business. With this in hand one can just check this equality.

[[Unfortunately we were now 10 minutes passed when the course should end. Realizing this Ribet brought everything to a close. As is the tradition at the end of a course the room erupted in applause.]]

Bibliography

[1] N. Katz, *Antwerp Proceedings*, Lecture Notes in Mathematics, vol. 350

[2] N. Boston, H.W. Lenstra, Jr., K.A. Ribet, *Quotients of group rings arising from two-dimensional representations*, C.R. Acad. Sci. Paris, t. **312**, Ser. I, Fev. 1991, p. 323–328

[3] C. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associate Algebras*, John Wiley and Sons, 1962

[4] H. Darmon, F. Diamond, R. Taylor, *Fermat's Last Theorem*, CMS Proceedings, 1996

[5] R. Coleman and B. Edixhoven, *Semi-simplicity of the U_p operator on wieght 2 modular forms*, to appear, 1996

[6] B. Edixhoven, *The weight in Serre's conjectures on modular forms*, Invent. Math. 109 (1992), no. 3, 563–594

[7] R. Hartshorne, *Algebraic Geometry*, GTM 52, 1977

[8] J. Igusa, *Fibre systems of Jacobian varieties*, I. Amer. J. of Maths., 78, 1956, p. 171-199; II, *id.*, p. 745-760; III, *id.*, 81, 1959, p. 453–476

[9] Katz and Mazur, *Arithmetic Moduli of Elliptic Curves*, Princeton University Press, 1985

[10] S. Lang, *Algebraic Number Theory*, GTM 110, Springer-Verlag, 2nd edition, 1994

[11] S. Lang, *Introduction to Modular Forms*, Grundlehren 222, Springer-Verlag, 1976

[12] H. Lenstra, B. de Smit, *Explicit construction of universal deformation rings*, Proceedings of the Conference at Boston University, Spring-Verlag, 1996

[13] *Lecture Notes in Mathematics, Vol. 349*

[14] B. Mazur, *Modular curves and the Eisenstein ideal*, Publ. Math. IHES **47** (1977) 133–186.

[15] B. Mazur, *Deforming Galois representations*, in: Galois groups over **Q**, Y. Ihara, K. Ribet, J-P. Serre, eds., MSRI Publ. **16**, Springer-Verlag, New York, Berlin, Heidelberg, 1989, pp. 385-437.

[16] J.S. Milne, *Abelian Varieties*, in *Arithmetic Geometry*, ed. Cornell and Silverman, Springer-Verlag, 1986

[17] D. Mumford, *Geometric Invariant Theory*

[18] D. Mumford, *Abelian Varieties*

[19] V. Kumar Murty, *Introduction to Abelian Varieties*, American Mathematical Society, 1993

[20] C. Queen, paper in *Number Theory and Algebra*, papers submitted to the J. of Numb. Theory, 1977

[21] K. Ribet, *Lecture Notes in Mathematics, Volume 601*, Springer-Verlag, 19xx

[22] K. A. Ribet, *Report on mod ℓ representations of $Gal(\overline{Q}/Q)$* Proceedings of Symposia in Pure Mathematics, Vol 55 (1994), Part 2

[23] J.P. Serre, *Algebraic Groups and Class Fields*, GTM 117, Springer-Verlag, 1988

[24] J.P. Serre, *A Course in Arithmetic*, GTM 7, Springer-Verlag, 1973

[25] J.P. Serre, *Local Fields*, GTM 67, Springer-Verlag, 1979

[26] J.P. Serre, *λ-adic Representations and Elliptic Curves*,

[27] J. P. Serre, *Sur les représentations modulaires de degré 2 de $Gal(\overline{Q}/Q)$*, Duke Mathematical Journal, Vol. 54, No. 1 (1987)

[28] J. P. Serre, *Propriétés galoisennes de points d'ordre fini des courbes elliptiques*, Invent. Math. 15, (1972), 259–331

[29] J.P. Serre, J. Tate, *Good Reduction of Abelian Varieties*, Annals of Mathematics, Vol. 88, No. 3, November, 1968, pp. 492–517

[30] O.F.G. Schilling, *Arithmetical Algebraic Geometry*, Proceedings of a Conference Held at Purdue University, Harper's, 1963

[31] G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions*, Princeton University Press, 1994

[32] G. Shimura, *Sur les intégrales attachées aux formes automorphes*, J. Math. Soc. Japan, 11, No. 4, 1959, pp. 291–311

[33] J. Silverman, *The Arithmetic of Elliptic Curves*, GTM 106, Springer-Verlag, 1986

[34] J. Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves*, GTM 151, Springer-Verlag, 1994.

[35] R. Taylor, A. Wiles, *Ring theoretic properties of certain Hecke algebras*, Annals of Math. **141** (1995), 553–572.

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

Index

$\Gamma(N)$-structures, 23, 24

Albanese, 50, 52–54, 69

Brauer-Nesbit theorem, 68
Buzzard, 45, 73, 89, 92, 107, 111, 125

complete intersection, 73, 127–129
congruence subgroup, 2
cusp forms, 2

deformations, 109, 111, 118, 119, 124–127
Deligne, 3
descent, 44, 75

Eichler, 3
Eichler-Shimura, 25, 51, 53–55, 57, 64, 71
eigenforms, 17
Eisenstein series, 15, 19, 40, 98

finite flat representations, 111, 112
finite flat schemes, 75–77, 103, 104, 110, 111

Galois representations, 1, 7, 34
Gorenstein, 67, 69, 72
Grothendieck, 75, 76

Hecke operator, 2, 11

inner product, 16, 20, 27

Jacobian, 1, 26, 50–54, 56, 74, 75, 77, 130

Lenstra, 125, 126, 130

Mazur, 6, 70–72, 75, 105, 106
modular curves, 1, 23, 41, 45, 47, 84

newform, 31
non-Gorenstein, 73

Ogus, 50, 75, 94, 97, 104, 111, 124, 130
optic, 69, 91, 127

Selmer group, 129
semistable, 103, 105–109, 111, 115, 122, 126
Serre, 3, 5, 7–9, 15, 21, 36, 40, 44, 83, 86, 87, 91–96, 98, 100–102
Serre's conjecture, 8
Shimura, 2, 3, 5, 26, 29, 31, 34, 52, 107
Shimura variety, 46

Tate curve, 65, 89, 104, 112