Nigel Boston

University of Wisconsin - Madison

THE PROOF OF FERMAT'S LAST THEOREM

Spring 2003

INTRODUCTION.

This book will describe the recent proof of Fermat's Last Theorem by Andrew Wiles, aided by Richard Taylor, for graduate students and faculty with a reasonably broad background in algebra. It is hard to give precise prerequisites but a first course in graduate algebra, covering basic groups, rings, and fields together with a passing acquaintance with number rings and varieties should suffice. Algebraic number theory (or arithmetical geometry, as the subject is more commonly called these days) has the habit of taking last year's major result and making it background taken for granted in this year's work. Peeling back the layers can lead to a maze of results stretching back over the decades.

I attended Wiles' three groundbreaking lectures, in June 1993, at the Isaac Newton Institute in Cambridge, UK. After returning to the US, I attempted to give a seminar on the proof to interested students and faculty at the University of Illinois, Urbana-Champaign. Endeavoring to be complete required several lectures early on regarding the existence of a model over $\mathbf{Q}$ for the modular curve $X_0(N)$ with good reduction at primes not dividing N. This work hinged on earlier work of Zariski from the 1950's. The audience, keen to learn new material, did not appreciate lingering over such details and dwindled rapidly in numbers.

Since then, I have taught the proof in two courses at UIUC, a two-week summer workshop at UIUC (with the help of Chris Skinner of the University of Michigan), and most recently a course in spring 2003 at the University of Wisconsin - Madison. To avoid getting bogged down as in the above seminar, it is necessary to assume some background. In these cases, references will be provided so that the interested students can fill in details for themselves. The aim of this work is to convey the strong and simple line of logic on which the proof rests. It is certainly well within the ability of most graduate students to appreciate the way the building blocks of the proof go together to give the result, even though those blocks may themselves be hard to penetrate. If anything, this book should serve as an inspiration for students to see why the tools of modern arithmetical geometry are valuable and to seek to learn more about them.

An interested reader wanting a simple overview of the proof should consult Gouvea [13], Ribet [25], Rubin and Silverberg [26], or my article [1]. A much more detailed overview of the proof is the one given by Darmon, Diamond, and Taylor [6], and the Boston conference volume [5] contains much useful elaboration on ideas used in the proof. The Seminaire Bourbaki article by Oesterlé and Serre [22] is also very enlightening. Of course, one should not overlook the original proof itself [38], [34] .

CONTENTS.

1

History and Overview

It is well-known that there are many solutions in integers to $x^2 + y^2 = z^2$, for instance $(3, 4, 5), (5, 12, 13)$. The Babylonians were aware of the solution $(4961, 6480, 8161)$ as early as around 1500 B.C. Around 1637, Pierre de Fermat wrote a note in the margin of his copy of Diophantus' *Arithmetica* stating that $x^n + y^n = z^n$ has no solutions in positive integers if $n > 2$. We will denote this statement for n $(FLT)_n$. He claimed to have a remarkable proof. There is some doubt about this for various reasons. First, this remark was published without his consent, in fact by his son after his death. Second, in his later correspondence, Fermat discusses the cases $n = 3, 4$ with no reference to this purported proof. It seems likely then that this was an off-the-cuff comment that Fermat simply omitted to erase. Of course $(FLT)_n$ implies $(FLT)_{\alpha n}$, for α any positive integer, and so it suffices to prove $(FLT)_4$ and $(FLT)_\ell$ for each prime number $\ell > 2$.

1.1 Proof of $(FLT)_4$ by Fermat

First, we must deal with the equation $x^2 + y^2 = z^2$. We may assume x, y, and z are positive and relatively prime (since otherwise we may divide out any common factors because the equation is homogeneous), and we see that one of x or y is even (since otherwise $z^2 \equiv 2 \pmod 4$, which is a contradiction). Suppose that x is even. Then

$$\left(\frac{z-y}{2}\right)\left(\frac{z+y}{2}\right) = \left(\frac{x}{2}\right)^2$$

with relatively prime factors on the left hand side and a square on the right hand side. Hence

$$\frac{z-y}{2} = b^2, \qquad \frac{z+y}{2} = a^2,$$

with $a, b \in \mathbf{Z}^+$. Then $y = a^2 - b^2$, $z = a^2 + b^2$, and $x = 2ab$.

[Alternatively, if $x^2 + y^2 = z^2$, then $(x + iy)/z$ has norm 1, and so by Hilbert's Theorem 90,

$$\frac{x+iy}{z} = \frac{a+ib}{a-ib} = \frac{(a^2-b^2) + i\,2ab}{a^2+b^2},$$

which yields the same result.]

Theorem 1.1 *$x^4 + y^4 = z^2$ has no solutions with x, y, z all nonzero, relatively prime integers.*

This implies $(FLT)_4$.

Proof: Say

$$\begin{aligned} x^2 &= 2ab \\ y^2 &= a^2 - b^2 \\ z &= a^2 + b^2 \end{aligned}$$

with a and b relatively prime. Clearly, b is even (y is odd, since x is even), and from $a^2 = y^2 + b^2$ we get

$$\begin{aligned} b &= 2cd \\ y &= c^2 - d^2 \\ a &= c^2 + d^2 \end{aligned}$$

with c and d relatively prime. Hence

$$x^2 = 2ab = 4cd(c^2 + d^2)$$

with c, d, and $c^2 + d^2$ relatively prime. Then

$$\begin{aligned} c &= e^2 \\ d &= f^2 \\ c^2 + d^2 &= g^2 \end{aligned}$$

whence $e^4 + f^4 = g^2$.

Note, however, that $z > a^2 = (g^2)^2 > g$, and so we are done by infinite descent (repeated application produces an infinite sequence of solutions with ever smaller positive integer z, a contradiction). QED

1.2 Proof of $(FLT)_3$

The first complete proof of this case was given by Karl Gauss. Leonhard Euler's proof from 1753 was quite different and at one stage depends on a fact that Euler did not justify (though it would have been within his knowledge to do so). We outline the proof - details may be found in [16], p. 285, or [23], p. 43.

Gauss's proof leads to a strategy that succeeds for certain other values of n too. We work in the ring $A = \mathbf{Z}[\zeta] = \{a + b\zeta : a, b \in \mathbf{Z}\}$, where ζ is a primitive cube root of unity. The key fact here is that A is a PID and hence a UFD. We also repeatedly use the fact that the units of A are precisely $\pm\zeta^i$ $(i = 0, 1, 2)$.

Theorem 1.2 *$x^3 + y^3 = uz^3$ has no solutions with $x, y, z \in A$, u a unit in A, $xyz \neq 0$.*

This certainly implies $(FLT)_3$.

Proof: By homogeneity, we may assume that x, y, z are relatively prime. Factoring $x^3 + y^3 = uz^3$ gives

$$(x + y)(x + \zeta y)(x + \zeta^2 y) = uz^3,$$

where the gcd of any 2 factors on the left divides $\lambda := 1 - \zeta$. If each gcd is 1, then each factor is a cube up to a unit. In any case, λ is "small" in that $|A/(\lambda)| = 3$. In particular, each element of A is either $0, \pm 1 \bmod \lambda$.

Lemma 1.3 *There are no solutions when $\lambda \not| xyz$.*

Proof: If $x \equiv 1 \pmod{\lambda}$, say $x = 1 + \lambda\alpha$, then

$$\begin{aligned} x^3 - 1 &= \lambda^3\alpha(1+\alpha)(\alpha - \zeta^2) \\ &\equiv \lambda^3\alpha(1+\alpha)(\alpha - 1) \pmod{\lambda^4} \\ &\equiv 0 \pmod{\lambda^4}, \end{aligned}$$

Plugging back in the equation, $(\pm 1) + (\pm 1) \equiv \pm u \pmod{\lambda^4}$, impossible since none of the 6 units u in A are 0 or ± 2 mod λ^4). QED

Lemma 1.4 *Suppose $\lambda \not| xy$, $\lambda | z$. Then $\lambda^2 | z$.*

Proof: Consider again $\pm 1 \pm 1 \equiv uz^3 \pmod{\lambda^4}$. If the left side is 0, then $\lambda^4 | z^2$, so $\lambda^2 | z$. If the left side is ± 2, then $\lambda | 2$, contradicting $|A/(\lambda)| = 3$. QED

Lemma 1.5 *Suppose that $\lambda \nmid xy$, $\lambda^k || z$, $k \geq 2$. Then there exists a solution with $\lambda \nmid xy$, $\lambda^{k-1} || z$.*

Proof: In this case, the gcd of any 2 factors on the left is λ. Hence we can assume that

$$\begin{aligned} (1) x + y &= u_1 \alpha^3 \lambda^t \\ (2) x + \zeta y &= u_2 \beta^3 \lambda \\ (3) x + \zeta^2 y &= u_3 \gamma^3 \lambda, \end{aligned}$$

where u_1, u_2, and u_3 are units, $t = 3k - 2$, and $\lambda \nmid \alpha, \beta, \gamma$. (1)+$\zeta$(2)+$\zeta^2$(2) yields (setting $x_1 = \beta$, $y_1 = \gamma$, and $z_1 = \alpha\lambda^{k-1}$)

$$x_1^3 + \epsilon_1 y_1^3 = \epsilon_2 z_1^3$$

with ϵ_1, ϵ_2 units. Reducing mod λ^2, we get

$$\pm 1 \pm \epsilon_1 \equiv 0,$$

which implies that $\epsilon_1 = \pm 1$. Replacing y_1 by $-y_1$ if necessary, we get

$$x_1^3 + y_1^3 = \epsilon_2 z^3.$$

QED

Finally, to prove the theorem, if $\lambda \nmid xyz$, we use lemma 1.2. If $\lambda \nmid xy$ but $\lambda | z$, we use lemmas 1.3 and 1.4. If $\lambda | x$ then $\lambda \nmid yz$, hence mod λ^3

$$0 \pm 1 \equiv \pm u$$

which implies that $u \equiv \pm 1 \pmod{\lambda^3}$, and hence $u = \pm 1$. Rearranging yields

$$(\pm z)^3 + (-y)^3 = x^3,$$

a case which has already been treated. QED

1.3 Further Efforts at Proof

Peter Dirichlet and Adrien Legendre proved $(FLT)_5$ around 1825, and Gabriel Lamé proved $(FLT)_7$ around 1839. If we set $\zeta = e^{2\pi i/\ell}$ (ℓ prime), and

$$\mathbf{Z}[\zeta] = \{a_0 + a_1\zeta + \ldots + a_{l-2}\zeta^{l-2} : a_i \in \mathbf{Z}\},$$

then there are cases when $\mathbf{Z}[\zeta]$ is not a UFD and the factorization method used above fails. (In fact, $\mathbf{Z}[\zeta]$ is a UFD if and only if $\ell \leq 19$.)

It turns out that the method can be resuscitated under weaker conditions. In 1844 Ernst Kummer began studying the ideal class group of $\mathbf{Q}(\zeta)$, which is a finite group that measures how far $\mathbf{Z}[\zeta]$ is from being a UFD [33]. Between 1847 and 1853, he published some masterful papers, which established almost the best possible result along these lines and were only really bettered by the recent approach detailed below, which began over 100 years later. In these papers, Kummer defined regular primes and proved the following theorem, where $h(\mathbf{Q}(\zeta))$ denotes the order of the ideal class group.

Definition 1.6 *Call a prime ℓ regular if $\ell \not| h(\mathbf{Q}(\zeta))$ (where $\zeta = e^{2\pi i/\ell}$). Otherwise, ℓ is called irregular.*

Remark 1.7 *The first irregular prime is 37 and there are infinitely many irregular primes. It is not known if there are infinitely many regular primes, but conjecturally this is so.*

Theorem 1.8 *(Kummer) (i) $(FLT)_\ell$ holds if ℓ is regular.*

(2) ℓ is regular if and only if ℓ does not divide the numerator of B_i for any even $2 \leq i \leq \ell - 3$.

Here B_n are the Bernoulli numbers defined by

$$\frac{x}{e^x - 1} = \sum (B_n/n!)x^n.$$

For instance, the fact that $B_{12} = -\frac{691}{2730}$ shows that 691 is irregular. We shall see the number 691 appearing in many different places.

Here the study of FLT is divided into two cases. The first case involves showing that there is no solution with $\ell \not| xyz$. The idea is to factor $x^\ell + y^\ell = z^\ell$ as

$$(x+y)(x+\zeta y)\cdots(x+\zeta^{\ell-1}y) = z^\ell,$$

where $\zeta = e^{2\pi i/\ell}$. The ideals generated by the factors on the left side are pairwise relatively prime by the assumption that $\ell \not| xyz$ (since $\lambda := 1 - \zeta$ has norm ℓ - compare the proof of $(FLT)_3$), whence each factor generates an ℓth power in the ideal class group of $\mathbf{Q}(\zeta)$. The regularity assumption then shows that these factors are principal ideals. We also use that any for unit u in $\mathbf{Z}[\zeta]$, $\zeta^s u$ is real for some $s \in \mathbf{Z}$. See [33] or [16] for more details.

The second case involves showing that there is no solution to FLT for $\ell | xyz$.

In 1823, Sophie Germain found a simple proof that if ℓ is a prime with $2\ell + 1$ a prime then the first case of $(FLT)_\ell$ holds. Arthur Wieferich proved in 1909 that if ℓ is a prime with $2^{\ell-1} \not\equiv 1 \pmod{\ell^2}$ then the first case of $(FLT)_\ell$ holds. Examples of ℓ that fail this are rare - the only known examples are 1093 and 3511. Moreover, similar criteria are known if $p^{\ell-1} \not\equiv 1$ (mod ℓ^2) and p is any prime ≤ 89 [15]. This allows one to prove the first case of $(FLT)_\ell$ for many ℓ.

Before Andrew Wiles, $(FLT)_\ell$ was known for all primes $2 <$

$\ell < 4 \times 10^6$ [3]; the method was to check that the conjecture of Vandiver (actually originating with Kummer and a refinement of his method) that $\ell \not| h(\mathbf{Q}(\zeta + \zeta^{-1}))$ holds for these primes. See [36]. The first case of $(FLT)_\ell$ was known for all primes $2 < \ell < 8.7 \times 10^{20}$.

1.4 Modern Methods of Proof

In 1916, Srinivasa Ramanujan proved the following. Let

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

Then

$$\tau(n) \equiv \sigma_{11}(n) \pmod{691},$$

where $\sigma_k(n) = \sum_{d|n} d^k$.

Δ is a modular form; this means that, if we set $q = e^{2\pi i z}$, Δ satisfies (among other conditions) $\Delta\left(\frac{az+b}{cz+d}\right) = (cz+d)^k \Delta(z)$ for all z in the upper half-plane $\mathrm{Im}(z) > 0$ and all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ with, in this case, ("weight") $k = 12$ and $\Gamma = SL_2(\mathbf{Z})$ (in general, we define a "level" N by having Γ defined as the group of matrices in $SL_2(\mathbf{Z})$ such that $N|c$; here $N = 1$). For instance, setting $a, b, d = 1, c = 0$, $\Delta(z+1) = \Delta(z)$, and this is why Δ can be written as a Fourier series in $q = e^{2\pi i z}$.

Due to work of André Weil in the 1940's and John Tate in the 1950's, the study of elliptic curves, that is curves of the form $y^2 = g(x)$, where g is a cubic with distinct roots, led to the study of Galois representations, i.e. continuous homomorphisms $Gal(\bar{\mathbf{Q}}/\mathbf{Q}) \to GL_2(R)$, where R is a complete local ring such as the finite field $\mathbf{F}_\ell$ or the ring of ℓ-adic integers $\mathbf{Z}_\ell$. In particular, given elliptic curve E defined over $\mathbf{Q}$ (meaning the coefficients of g are in $\mathbf{Q}$), and any rational prime ℓ, there exist associated Galois representations $\rho_{\ell,E} : Gal(\bar{\mathbf{Q}}/\mathbf{Q}) \to GL_2(\mathbf{Z}_\ell)$ and (by reduction mod ℓ) $\overline{\rho}_{\ell,E} : Gal(\bar{\mathbf{Q}}/\mathbf{Q}) \to GL_2(\mathbf{F}_\ell)$. These encode much information about the curve.

A conjecture of Jean-Pierre Serre associates to a certain kind of modular form f (cuspidal eigenforms) and to a rational prime ℓ a Galois representation, $\rho_{\ell,f}$. All known congruences for τ follow from a systematic study of the representations associated to Δ. This conjecture was proved by Pierre Deligne [7] (but note that he really only wrote the details for Δ - extensive notes of Brian Conrad http://www.math.lsa.umich.edu/$\sim$ bdconrad/bc.ps can be used to fill in details here) in 1969 for weights $k > 2$. For $k = 2$ it follows from earlier work of Martin Eichler and Goro Shimura [31]. For $k = 1$ it was later established by Deligne and Serre [8].

These representations $\rho_{\ell,f}$ share many similarities with the representations $\rho_{\ell,E}$. Formalizing this, a conjecture of Yutaka Taniyama of 1955, later put on a solid footing by Shimura, would attach a modular form of this kind to each elliptic curve over $\mathbf{Q}$. Thus, we have the following picture

$$\begin{tikzcd} & \{Repns\ from\ elliptic\ curves\} \arrow[d, phantom, sloped, "\subseteq"] \\ \{Repns\ from\ certain\ modular\ forms\} \arrow[r, phantom, "\subseteq"] & \{Admissible\ Galois\ representat \end{tikzcd}$$

In 1985, Gerhard Frey presented a link with FLT. If we assume that a, b, c are positive integers with $a^\ell + b^\ell = c^\ell$, and consider the elliptic curve $y^2 = x(x - a^\ell)(x + b^\ell)$ (called a *Frey curve*), this curve is unlikely to be modular, in the sense that

$\overline{\rho}_{\ell,E}$ turns out to have properties that a representation associated to a modular form should not.

The *Shimura-Taniyama conjecture*, however, states that any given elliptic curve is modular. That is, given E, defined over $\mathbf{Q}$, we consider its L-function $L(E,s) = \sum a_n/n^s$. This conjecture states that $\sum a_n q^n$ is a modular form. Equivalently, every $\rho_{\ell,E}$ is a $\rho_{\ell,f}$ for some modular form f.

In 1986, Kenneth Ribet (building on ideas of Barry Mazur) showed that these Frey curves are definitely not modular. His strategy was to show that if the Frey curve is associated to a modular form, then it is associated to one of weight 2 and level 2. No cuspidal eigenforms of this kind exist, giving the desired contradiction. Ribet's approach (completed by Fred Diamond and others) establishes in fact that the weak conjecture below implies the strong conjecture (the implication being the so-called ϵ-*conjecture*). The strong conjecture would imply many results - unfortunately, no way of tackling this is known.

Serre's weak conjecture [30] says that all Galois representations $\overline{\rho} : Gal(\bar{\mathbf{Q}}/\mathbf{Q}) \to GL_2(k)$ with k a finite field, and such that $\det(\overline{\rho}(\tau)) = -1$, where τ denotes a complex conjugation, (this condition is the definition of $\overline{\rho}$ being *odd*) come from modular forms.

Serre's strong conjecture [30] states that ρ comes from a modular form of a particular type (k, N, ϵ) with k, N positive integers (the weight and level, met earlier) and $\epsilon : (\mathbf{Z}/N\mathbf{Z})^\times \to \mathbf{C}^\times$ (the Nebentypus). In the situations above, ϵ is trivial.

In 1986, Mazur found a way to parameterize certain collections of Galois representations by rings. Frey curves are semistable, meaning that they have certain mild singularities modulo primes. Wiles with Richard Taylor proved in 1994 that every semistable elliptic curve is modular.

In a picture we have (restricting to certain subsets to be defined later):

$$\begin{array}{ccc} & & \{Certain\ semistable\ elliptic\ curves\} \\ & & |\cap \\ \{Certain\ modular\ forms\} & \subseteq & \{Certain\ semistable\ Galois\ representations \end{array}$$

Wiles' idea is, first, following Mazur to parametrize the sets on the bottom line by local rings $\mathcal{T}$ and $\mathcal{R}$. The inclusion translates into a surjection from $\mathcal{R} \to \mathcal{T}$. Using some clever commutative algebra, Wiles obtains conditions for such a map to be an isomorphism. Using Galois cohomology and the theory of modular curves, it is checked that these conditions generally hold. The isomorphism of $\mathcal{R}$ and $\mathcal{T}$ translates back into the two sets on the bottom line being equal. It then follows that every semistable elliptic curve is modular.

In particular our particular Frey curves are modular, contradicting the conclusion of Ribet's work and establishing that counterexamples to Fermat's Last Theorem do not exist.

The Big Picture. An outline to the strategy of the proof has been given. A counterexample to Fermat's Last Theorem would yield an elliptic curve (Frey's curve) with remarkable properties. This curve is shown as follows not to exist. Associated to elliptic curves and to certain modular forms are Galois representations. These representations share some features, which might be used to define admissible representations. The aim is to show that all such admissible representations come from modular forms (and so in particular the elliptic curve ones do, implying that Frey's curves are modular, enough for a contradiction). We shall parametrize special subsets of Galois representations by complete Noetherian local rings and our aim will amount to showing that a given map between such rings is an isomorphism. This is achieved by some commutative algebra, which reduces the problem to computing some invariants, accomplished via Galois cohomology. The first step is to define (abstractly) Galois representations.

2
Profinite Groups and Complete Local Rings

2.1 Profinite Groups

Definition 2.1 *A directed set is a partially ordered set I such that for all $i, j \in I$ there is a $k \in I$ with $i \leq k$ and $j \leq k$.*

Example: Let G be a group. Index the normal subgroups of finite index by I. Say $i \geq j$ if $N_i \subseteq N_j$. If k corresponds to $N_k = N_i \cap N_j$ then $i, j \leq k$, so we have a directed set.

Definition 2.2 *An inverse system of groups is a collection of groups indexed by a directed set I, together with group homomorphisms $\pi_{ij} : G_i \to G_j$ whenever $i \geq j$. We insist that $\pi_{ii} = \mathrm{Id}$, and that $\pi_{jk}\pi_{ij} = \pi_{ik}$.*

Example: Index the normal subgroups of finite index by I as above. Setting $G_i = G/N_i$, and $\pi_{ij} : G_i \to G_j$ to be the natural quotient map whenever $i \geq j$, we get an inverse system of groups.

We now form a new category, whose objects are pairs $(H, \{\phi_i : i \in I\})$, where H is a group and each $\phi_i : H \to G_i$ is a group homomorphism, with the property that

$$\begin{tikzcd}
 & H \arrow[dl, "\phi_i"'] \arrow[dr, "\phi_j"] & \\
G_i \arrow[rr, "\pi_{ij}"'] & & G_j
\end{tikzcd}$$

commutes whenever $i \geq j$. Given two elements $(H, \{\phi_i\})$ and $(J, \{\psi_i\})$, we define a morphism between them to be a group homomorphism $\theta : H \to J$ such that

$$\begin{tikzcd}
H \arrow[rr, "\theta"] \arrow[dr, "\phi_i"'] & & J \arrow[dl, "\psi_i"] \\
 & G_i &
\end{tikzcd}$$

commutes for all $i \in I$.

Example: Continuing our earlier example, $(G, \{\phi_i\})$ is an object of the new category, where $\phi_i : G \to G/N_i$ is the natural quotient map.

Definition 2.3 $\varprojlim_{i \in I} G_i$ *is the terminal object in the new category, called the inverse limit of the* G_i*. That is,* $\varprojlim G_i$ *is the unique object* $(X, \{\chi_i\})$ *such that given any object* $(H, \{\phi_i\})$ *there is a unique morphism*

$$(H, \{\phi_i\}) \to (X, \{\chi_i\}).$$

The existence of a terminal object in this category will be proved below, after the next example.

Example: Continuing our earlier example, the group above, X, is the profinite completion $\hat{G}$ of G. Since $\hat{G}$ is terminal, there is a unique group homomorphism $G \to \hat{G}$. If this is an isomorphism then we say that G is *profinite* (or *complete*). For instance, it will be shown below that $Gal(\bar{\mathbf{Q}}/\mathbf{Q})$ is a profinite group.

We have the following commutative diagram for every object $(H, \{\phi_i\})$ of the new category and every $N_i \subseteq N_j$,

$$\begin{tikzcd}
H \arrow[rrrr, bend left, "\exists!"] \arrow[dr, "\phi_i"'] \arrow[drrr, "\phi_j"] & & G \arrow[rr, "\exists!"] \arrow[dl] \arrow[dr] & & \hat{G} \arrow[dlll] \arrow[dl] \\
& G/N_i \arrow[rr] & & G/N_j &
\end{tikzcd}$$

$\hat{G}$ contains all relevant information on finite quotients of G. $G \to \hat{G}$ is called the *profinite completion* of G. If we only use those finite quotients which are $\mathcal{C}$-groups, then we obtain the *pro-$\mathcal{C}$ completion* of G instead.

Exercise: Prove that $\hat{G} \to \hat{\hat{G}}$ is an isomorphism, so that $\hat{G}$ is profinite/complete.

We return to the general case and we now need to prove the existence of $\varprojlim_{i\in I} G_i$. To do this, let

$$C = \prod_{i\in I} G_i,$$

and $\pi_i : C \to G_i$ be the ith projection. Let $X = \{c \in C | \pi_{ij}(\pi_i(c)) = \pi_j(c)\ \forall i \geq j\}$. We claim that

$$\varprojlim_{i\in I} G_i = (X, \{\pi_i|_X\}).$$

Proof: (i) $(X, \{\pi_i|_X\})$ is an object in the new category, since X is a group (check!) and the following diagram commutes for all

$i \geq j$ (by construction)

$$\begin{tikzcd}
& X \arrow[dl, "\pi_i|_X"'] \arrow[dr, "\pi_j|_X"] & \\
G_i \arrow[rr, "\pi_{ij}"'] & & G_j
\end{tikzcd}$$

(ii) Given any $(H, \{\phi_i\})$ in the new category, define $\phi(h) = (\phi_i(h))_{i \in I}$, and check that this is a group homomorphism $\phi : H \to X$ such that

$$\begin{tikzcd}
H \arrow[rr, "\phi"] \arrow[dr, "\phi_i"'] & & X \arrow[dl, "\pi_i"] \\
& G_i &
\end{tikzcd}$$

and that ϕ is forced to be the unique such map. QED

Example: Let $G = \mathbf{Z}$ and let us describe $\hat{G} = \hat{\mathbf{Z}}$. The finite quotients of G are $G_i = \mathbf{Z}/i$, and $i \geq j$ means that $j|i$. Hence

$$\hat{\mathbf{Z}} = \{(a_1, a_2, a_3, \ldots) | a_i \in \mathbf{Z}/i \text{ and } a_i \equiv a_j \pmod{j} \text{ whenever } j|i\}.$$

Then for $a \in \mathbf{Z}$, the map $a \mapsto (a, a, a \ldots) \in \hat{\mathbf{Z}}$ is a homomorphism of $\mathbf{Z}$ into $\hat{\mathbf{Z}}$.

Now consider $\bar{\mathbf{F}}_p = \cup_n \mathbf{F}_{p^n}$. Then, if $m|n$,

$$\begin{tikzcd}
Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p) \arrow[r, "{\text{restriction}, \phi_n}"] \arrow[dr, "{\text{restriction}, \phi_m}"'] & Gal(\mathbf{F}_{p^n}/\mathbf{F}_p) \cong \mathbf{Z}/n \arrow[d, "\pi_{nm}"] \\
& Gal(\mathbf{F}_{p^m}/\mathbf{F}_p) \cong \mathbf{Z}/m
\end{tikzcd}$$

Note that $Gal(\mathbf{F}_{p^n}/\mathbf{F}_p)$ is generated by the *Frobenius automorphism* $Fr : x \mapsto x^p$. We see that $(Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p), \{\phi_n\})$ is an object in the new category corresponding to the inverse system. Thus there is a map $Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p) \to \hat{\mathbf{Z}}$.

We claim that this map is an isomorphism, so that $Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p)$ is profinite. This follows from our next result.

Theorem 2.4 *Let L/K be a (possibly infinite) separable, algebraic Galois extension. Then $Gal(L/K) \cong \varprojlim Gal(L_i/K)$, where the limit runs over all finite Galois subextensions L_i/K.*

Proof: We have restriction maps:

$$\begin{tikzcd}Gal(L/K) \arrow[r, "\phi_i"] \arrow[dr, "\phi_j"'] & Gal(L_i/K) \arrow[d] \\ & Gal(L_j/K)\end{tikzcd}$$

whenever $L_j \subseteq L_i$, i.e. $i \geq j$. We use the projection maps to form an inverse system, so, as before, $(Gal(L/K), \{\phi\})$ is an object of the new category and we get a group homomorphism

$$Gal(L/K) \xrightarrow{\phi} \varprojlim Gal(L_i/K).$$

We claim that ϕ is an isomorphism.

(i) Suppose $1 \neq g \in Gal(L/K)$. Then there is some $x \in L$ such that $g(x) \neq x$. Let L_i be the Galois (normal) closure of $K(x)$. This is a finite Galois extension of K, and $1 \neq g|_{L_i} = \phi_i(g)$, which yields that $1 \neq \phi(g)$. Hence ϕ is injective.

(ii) Take $(g_i) \in \varprojlim Gal(L_i/K)$ - this means that $L_j \subseteq L_i \Rightarrow g_i|L_j = g_j$. Then define $g \in Gal(L/K)$ by $g(x) = g_i(x)$ whenever $x \in L_i$. This is a well-defined field automorphism and $\phi(g) = (g_i)$. Thus ϕ is surjective. QED

For the rest of this section, we assume that the groups G_i are all finite (as, for example, in our running example). Endow the finite G_i in our inverse system with the discrete topology. G_i is certainly a totally disconnected Hausdorff space. Since these properties are preserved under taking products and subspaces, $\varprojlim G_i \subseteq \prod G_i$ is Hausdorff and totally disconnected as well. Furthermore $\prod G_i$ is compact by Tychonoff's theorem.

Exercise: If $f, g : A \to B$ are continuous (A, B topological spaces) with A, B Hausdorff, then $\{x | f(x) = g(x)\}$ is closed. Deduce that

$$\varprojlim G_i = \bigcap_{i \geq j} \left\{ c \in \prod G_i : \pi_{ij}(\pi_i(c)) = \pi_j(c) \right\}$$

is closed in $\prod G_i$, therefore is compact. In summary, $\varprojlim G_i$ is a compact, Hausdorff, totally disconnected topological space.

Exercise: The natural inclusion $\mathbf{Z} \to \hat{\mathbf{Z}}$ maps $\mathbf{Z}$ onto a dense subgroup. In fact, for any group G, its image in $\hat{G}$ is dense, but the kernel of $G \to \hat{G}$ need not be trivial. This happens if and only if G is residually finite (meaning that the intersection of all its subgroups of finite index is trivial).

If we denote by Fr the element of $Gal(\bar{\mathbf{F}}_q/\mathbf{F}_q)$ given by $Fr(x) = x^q$, i.e. the Frobenius automorphism, then Fr does not generate the Galois group, but the group which it does generate is dense (by the last exercise), and so we say that $Gal(\bar{\mathbf{F}}_q/\mathbf{F}_q)$ is *topologically finitely generated* by one element Fr (and so is *procyclic*).

2.2 Complete Local Rings

We now carry out the same procedure with rings rather than groups and so define certain completions of them. Let R be a commutative ring with identity 1, I any ideal of R. For $i \geq j$ we have a natural quotient map

$$R/I^i \xrightarrow{\pi_{ij}} R/I^j.$$

These rings and maps form an inverse system (now of rings). Proceeding as in the previous section, we can form a new category. Then the same proof gives that there is a unique terminal object, $R_I = \varprojlim_i R/I^i$, which is now a ring, together with a unique ring homomorphism $R \to R_I$, such that the following diagram commutes:

$$\begin{tikzcd}
R \arrow[rrr] \arrow[dr, "\pi_i"] \arrow[drr, "\pi_j"] & & & R_I \arrow[dll] \arrow[dl] \\
& R/I^i \arrow[r, "\pi_{ij}"'] & R/I^j &
\end{tikzcd}$$

Note that R_I depends on the ideal chosen. It is called the *I-adic completion* of R (do not confuse it with the localization of R at I). We call R (I-adically) *complete* if the map $R \to R_I$ is an isomorphism. Then R_I is complete. If I is a maximal ideal $\mathbf{m}$, then we check that R_I is *local*, i.e. has a unique maximal ideal, namely $\mathbf{m}\hat{R}$. This will be proven for the most important example, $\mathbf{Z}_p$ (see below), at the start of the next chapter.

Example: If $R = \mathbf{Z}$, $\mathbf{m} = p\mathbf{Z}$, p prime, then $R_I \cong \mathbf{Z}_p$, the *p-adic integers*. The additive group of this ring is exactly the pro-p completion of $\mathbf{Z}$ as a group. In fact,

$$\mathbf{Z}_p = \{(a_1, a_2, \ldots) : a_i \in \mathbf{Z}/p^i, a_i \equiv a_j \pmod{p^j} \text{ if } i \geq j\}.$$

Note that $\hat{\mathbf{Z}}$ will always be used to mean the profinite (rather than any I-adic) completion of $\mathbf{Z}$.

Exercise: Let $R = \mathbf{Z}$ and $I = 6\mathbf{Z}$. Show that the I-adic completion of $\mathbf{Z}$ is isomorphic to $\mathbf{Z}_2 \times \mathbf{Z}_3$ and so is not local (it's called semilocal).

Exercise: Show that $\hat{Z} = \prod_\ell \mathbf{Z}_p$, the product being over all rational primes.

Exercise: Show that the ideals of $\mathbf{Z}_p$ are precisely $\{0\}$ and $p^i\mathbf{Z}_\ell$

$(i \geq 0)$. (This also follows from the theory developed in chapter 3.)

We will be interested in the category $\mathcal{C}$ whose objects are the complete local Noetherian rings with a given finite residue field (that is, the ring modulo its maximal ideal) k. In this category a morphism is required to make the following diagram commute:

$$\begin{tikzcd} R \arrow[r, "\phi"] \arrow[d] & S \arrow[d] \\ k \arrow[r, "\mathrm{Id}"'] & k \end{tikzcd}$$

where the vertical maps are the natural projections. This is equivalent to requiring that $\phi(\mathbf{m}_R) \subseteq \mathbf{m}_S$, where $\mathbf{m}_R$ (respectively $\mathbf{m}_S$) is the maximal ideal of R (respectively S).

As an example, if $k = \mathbf{F}_\ell$, then $\mathbf{Z}_\ell$ is an object of $\mathcal{C}$. By a theorem of Cohen [2], the objects of $\mathcal{C}$ are of the form

$$W(k)[[T_1, \ldots, T_m]]/(\text{ideal}),$$

where $W(k)$ is a ring called the ring of infinite Witt vectors over k (see chapter 3 for an explicit description of it). Thus $W(k)$ is the initial object of $\mathcal{C}$. In the case of $\mathbf{F}_p$, $W(\mathbf{F}_p) = \mathbf{Z}_p$.

Exercise: If R is a ring that is I-adically complete, then $GL_n(R) \cong \varprojlim_i GL_n(R/I^i)$ (the maps $GL_n(R/I^i) \to GL_n(R/I^j)$ being the natural ones).

Note that the topology on R induces the product topology on $M_n(R)$ and thence the subspace topology on $GL_n(R)$.

The Big Picture. We shall seek to use continuous group homomorphisms (*Galois representations*)

$$Gal(\bar{\mathbf{Q}}/\mathbf{Q}) \to GL_n(R),$$

where R is in some $\mathcal{C}$, to parametrize the homomorphisms that elliptic curves and modular forms naturally produce. In this chapter we have constructed these groups and rings and explained their topologies. Next, we study the internal structure of both sides, notably certain important subgroups of the left side. This will give us the means to characterize Galois representations in terms of their effect on these subgroups.

3

Infinite Galois Groups: Internal Structure

We begin with a short investigation of $\mathbf{Z}_p$. A good reference for this chapter is [28].

We first check that $p^n\mathbf{Z}_p$ is the kernel of the map $\mathbf{Z}_p \to \mathbf{Z}/p^n$. Hence, we have

$$\mathbf{Z}_p \supset p\mathbf{Z}_p \supset p^2\mathbf{Z}_p \ldots$$

If $x \in p^n\mathbf{Z}_p - p^{n+1}\mathbf{Z}_p$, then we say that the valuation of x, $v(x) = n$. Set $v(0) = \infty$.

Exercise: x is a unit in $\mathbf{Z}_p$ if and only if $v(x) = 0$

Corollary 3.1 *Every $x \in \mathbf{Z}_p - \{0\}$ can be uniquely written as $p^{v(x)}u$ where u is a unit.*

In fact

$(*) : (1)v(xy) = v(x) + v(y), (2)v(x + y) \geq \min(v(x), v(y))$.

We define a metric on $\mathbf{Z}_p$ as follows: set $d(x, y) = c^{v(x-y)}$ for a fixed $0 < c < 1, x \neq y \in \mathbf{Z}_p$ ($d(x, x) = 0$). We have something stronger than the triangle inequality, namely $d(x, z) \leq$

$\max(d(x,y), d(y,z))$ for all $x, y, z \in \mathbf{Z}_p$. This has unusual consequences such as that every triangle is isosceles and every point in an open unit disc is its center.

The metric and profinite topologies then agree, since $p^n\mathbf{Z}_p$ is a base of open neighborhoods of 0 characterized by the property $v(x) \geq n \iff d(0,x) \leq c^n$.

By $(*)(1)$, $\mathbf{Z}_p$ is an integral domian. Its quotient field is called $\mathbf{Q}_p$, the field of p-adic numbers. We have the following diagram of inclusions.

$$\begin{tikzcd}
\bar{\mathbf{Q}} \arrow[r, hook] & \bar{\mathbf{Q}}_p \\
\mathbf{Q} \arrow[r, hook] \arrow[u, hook] & \mathbf{Q}_p \arrow[u, hook] \\
\mathbf{Z} \arrow[r, hook] \arrow[u, hook] & \mathbf{Z}_p \arrow[u, hook]
\end{tikzcd}$$

This then produces restriction maps

$$Gal(\bar{\mathbf{Q}}_p/\mathbf{Q}_p) \to Gal(\bar{\mathbf{Q}}/\mathbf{Q}). \tag*{$(*)$}$$

We can check that this is a continuous group homomorphism (defined up to conjugation only). Denote $Gal(\bar{K}/K)$ by G_K.

Definition 3.2 *Given a continuous group homomorphism $\rho : G_{\mathbf{Q}} \to GL_n(R)$, $(*)$ yields by composition a continuous group homomorphism $\rho_p : G_{\mathbf{Q}_p} \to GL_n(R)$. The collection of homomorphisms ρ_p, one for each rational prime p, is called the local data attached to ρ.*

The point is that $G_{\mathbf{Q}_p}$ is much better understood than $G_{\mathbf{Q}}$; in fact even presentations of $G_{\mathbf{Q}_p}$ are known, at least for $p \neq 2$ [18]. We next need some of the structure of $G_{\mathbf{Q}_p}$ and obtain this by investigating finite Galois extensions K of $\mathbf{Q}_p$ and how $Gal(K/\mathbf{Q}_p)$ acts.

Now,

$$\mathbf{Z}_p - \{0\} = \{p^n u | n \geq 0, u \text{ is a unit}\},$$

whence

$$\mathbf{Q}_p - \{0\} = \{p^n u | u \text{ is a unit in} \mathbf{Z}_p\}.$$

We can thus extend $v : \mathbf{Z}_p - \{0\} \to \mathbf{N}$ to a map $v : \mathbf{Q}_p^\times \to \mathbf{Z}$ which is a homomorphism of groups.

Definition 3.3 *A discrete valuation w on a field K is a surjective homomorphism $w : K^\times \to \mathbf{Z}$ such that*

$$w(x+y) \geq \min(w(x), w(y)$$

for all $x, y \in K$ (we take $w(0) = \infty$).

Example: The map $v : \mathbf{Q}_p \to \mathbf{Z}$ is a discrete valuation, since $(*)$ extends to $\mathbf{Q}_p$.

Exercise: If K has a discrete valuation then

$$A = \{x \in K | w(x) \geq 0\}$$

is a ring, and

$$\mathbf{m} = \{x \in K | w(x) > 0\}$$

is its unique maximal ideal. Choose π so that $w(\pi) = 1$. Every element $x \in K^\times$ can be uniquely written as $x = \pi^{w(x)} u$, where u is a unit in A.

Remark: $A, A/\mathbf{m}$, and π are called respectively the *valuation ring*, *the residue field*, and a *uniformizer* of w. As with the valuation on $\mathbf{Q}_p$, w yields a metric on K. An alternative way of describing the elements of A is by the power series $\{c_0 + c_1\pi + c_2\pi^2 + \ldots | c_i \in S\}$, where $S \subset A$ is chosen to contain exactly one element of each coset of $A/\mathbf{m}$. The connection with our approach is by mapping the typical element to $(c_0 \pmod{\pi}, c_0 + c_1\pi \pmod{\pi)^2}, c_0 + c_1\pi + c_2\pi^2 \pmod{\pi)^3}, \ldots) \in \prod A/\mathbf{m}^i = A$. The description extends to K by having $K = \{\sum_{i=N}^{\infty} c_i\pi^i\}$, i.e. Laurent series in π, coefficients in S.

Corollary 3.4 *The ideal* $\mathbf{m}$ *is equal to the principal ideal* (π)*, and all ideals of* A *are of the form* (π^n)*, so* A *is a PID.*

Let $K/\mathbf{Q}_p$ be a finite Galois extension, and define a norm $N : K^\times \to \mathbf{Q}_p^\times$ by

$$x \mapsto \prod_{\sigma \in G} \sigma(x)$$

where $G = Gal(K/\mathbf{Q}_p)$. The composition of homomorphisms

$$K^\times \xrightarrow{N} \mathbf{Q}_p^\times \xrightarrow{v} \mathbf{Z}$$

is nonzero with some image $f\mathbf{Z}$. We then define $w : K^\times \to \mathbf{Z}$ by $w = \frac{1}{f} v \circ N$. Then w is a discrete valuation on K. f is called the *residue degree* of K, and we say that w extends v with *ramification index* e if $w|_{\mathbf{Q}_p} = ev$. For any $x \in \mathbf{Q}_p$, we have

$$ev(x) = w(x) = \frac{1}{f}v(N(x)) = \frac{1}{f}v(x^n) = \frac{n}{f}v(x),$$

so that $ef = n = [K : \mathbf{Q}_p]$.

Proposition 3.5 *The discrete valuation* w *is the unique discrete valuation on* K *which extends* v*.*

Proof: A generalization of the proof that any two norms on a finite dimensional vector space over $\mathbf{C}$ are equivalent ([4],[29] Chap. II). QED

Exercise: Let A be the valuation ring of K, and $\mathbf{m}$ the maximal ideal of A. Prove that the order of $A/\mathbf{m}$ is p^f.

We have a collection of embeddings as follows:

$$\begin{tikzcd}
A \arrow[r, hook] & K \\
\mathbf{Z}_p \arrow[u, hook] \arrow[r, hook] & \mathbf{Q}_p \arrow[u, hook]
\end{tikzcd}$$

We note that the action of G on K satisfies $w(\sigma(x)) = w(x)$, for all $\sigma \in G$, $x \in K$, since w and $w \circ \sigma$ both extend v (or by using the explicit definition of w).

This property is crucial in establishing certain useful subgroups of G.

Definition 3.6 *The ith ramification subgroup of $G = Gal(K/\mathbf{Q}_p)$ is*

$$G_i = \{\sigma \in G | w(\sigma(x) - x) \geq i + 1 \text{ for all } x \in A\},$$

for $i = -1, 0, \ldots$. (See [29], Ch. IV.)

G_i is a group, since $1 \in G_i$, and if $\sigma, \tau \in G$, then

$$\begin{aligned}
w(\sigma\tau(x) - x) &= w(\sigma(\tau(x) - x) + (\sigma(x) - x)) \\
&\geq \min(w(\sigma(\tau(x) - x), w(\sigma(x) - x)) \\
&\geq i + 1,
\end{aligned}$$

so that $\sigma\tau \in G_i$. This is sufficient since G is finite. Moreover,

$$\begin{aligned}
\sigma\tau\sigma^{-1}(x) - x &= \sigma(\tau\sigma^{-1}(x) - \sigma^{-1}(x)) \\
&= \sigma(\tau(\sigma^{-1}(x)) - \sigma^{-1}(x))
\end{aligned}$$

shows that $G_i \lhd G$.

We have that

$$G = G_{-1} \supseteq G_0 \supseteq G_1 \supseteq \cdots,$$

where we call G_0 the *inertia subgroup* of G, and G_1 the *wild inertia subgroup* of G.

Exercise: Let π be a uniformizer of K. Show that in fact

$$G_i = \{\sigma \in G | w(\sigma(\pi) - \pi) \geq i + 1\}.$$

These normal subgroups determine a filtration of G, and we now study the factor groups in this filtration.

Theorem 3.7 *(a) The quotient G/G_0 is canonically isomorphic to $Gal(k/\mathbf{F}_p)$, with $k = A/\mathbf{m}$, hence it is cyclic of order f.*

(b) Let U_0 be the group of units of A. Then $U_i = 1 + (\pi^i)$ ($i \geq 1$) is a subgroup of U_0. For all $\sigma \in G$, the map $\sigma \mapsto \sigma(\pi)/\pi$ induces an injective group homomorphism $G_i/G_{i+1} \hookrightarrow U_i/U_{i+1}$.

Proof: (a) Let $\sigma \in G$. Then σ acts on A, and sends $\mathbf{m}$ to $\mathbf{m}$. Hence it acts on $A/\mathbf{m} = k$. This defines a map $\phi : G \to Gal(k/\mathbf{F}_p)$ by sending σ to the map $x + \mathbf{m} \mapsto \sigma(x) + \mathbf{m}$. We now examine the kernel of this map.

$$\begin{aligned} \ker \phi &= \{\sigma \in G | \sigma(x) - x \in \mathbf{m} \text{ for all } x \in A\} \\ &= \{\sigma \in G | w(\sigma(x) - x) \geq 1 \text{ for all } x \in A\} = G_0. \end{aligned}$$

This shows that ϕ induces an injective homomorphism from G/G_0 to $Gal(k/\mathbf{F}_p)$.

As for surjectivity, choose $a \in A$ such that the image $\bar{a}$ of a in k has $k = \mathbf{F}_p(\bar{a})$. Let

$$p(x) = \prod_{\sigma \in G} (x - \sigma(a)).$$

Then $p(x)$ is a monic polynomial with coefficients in A, and one root is a. Then

$$\overline{p(x)} = \prod_{\sigma \in G} (x - \overline{\sigma(a)}) \in k[x]$$

yields that all conjugates of $\bar{a}$ are of the form $\overline{\sigma(a)}$. For $\tau \in Gal(k/\mathbf{F}_p)$, $\tau(\bar{a})$ is such a conjugate, whence it is equal to some $\overline{\sigma(a)}$. Then the image of σ is τ.

(b) $\sigma \in G_i \iff w(\sigma(\pi) - \pi) \geq i + 1 \iff \sigma(\pi)/\pi \in U_i$.

This map is independent of choice of uniformizer. Suppose $\pi' = \pi u$ is another uniformizer, where u is a unit. Then

$$\frac{\sigma(\pi')}{\pi'} = \frac{\sigma(\pi)}{\pi}\frac{\sigma(u)}{u}.$$

but for $\sigma \in G_i$, we have that $i+1 \leq w(\sigma(u) - u) = w(\sigma(u)/u - 1) + w(u)$, so that $\sigma(u)/u \in U_{i+1}$, i.e. $\sigma(\pi')/\pi'$ differs from $\sigma(\pi)/\pi$ by an element of U_{i+1}.

The map θ_i is a homomorphism, since

$$\frac{\sigma\tau(\pi)}{\pi} = \frac{\sigma(\pi)}{\pi}\frac{\tau(\pi)}{\pi}\frac{\sigma(u)}{u},$$

where $u = \tau(\pi)/\pi$, and, as above, $\sigma(u)/u \in U_{i+1}$.

Finally the map θ_i is injective. To see this assume that $\sigma(\pi)/\pi \in U_{i+1}$. Then $\sigma(\pi) = \pi(1 + y)$ with $y \in (\pi^{i+1})$. Hence $\sigma(\pi) - \pi = \pi y$ has valuation at least $i + 2$, that is, $\sigma \in G_{i+1}$. QED

Proposition 3.8 *We have that U_0/U_1 is canonically isomorphic to $k^\times$, and is thus cyclic of order $p^f - 1$, and for $i \geq 1$, U_i/U_{i+1} embeds canonically into $\pi^i/\pi^{i+1} \cong k^+$, and so is elementary p-abelian.*

Proof: The map $x \mapsto x \pmod{\mathbf{m}}$ takes U_0 to $k^\times$, and is surjective. Its kernel is $\{x | x \equiv 1 \pmod{\mathbf{m}}\} = U_1$, whence $U_0/U_1 \cong k^\times$.

The map $1 + x \mapsto x \pmod{\pi}^{i+1}$ takes U_i to π^i/π^{i+1} and is surjective with kernel U_{i+1}. Moreover, since π acts trivially on π^i/π^{i+1}, this is an $(A/\mathbf{m})$-module (i.e. k-vector space). Its dimension is 1, since otherwise there would be an ideal of A strictly between π^i and π^{i+1}. QED

We note in particular that G is solvable, since the factors in its filtration are all abelian by the above. Thus, $G_{\mathbf{Q}_p}$ is prosolvable. Specifically, we have the following inclusions:

$$\begin{tikzcd}
G \\
G_0 \arrow[u, hook, "\text{cyclic}"] \\
G_1 \arrow[u, hook, "\text{cyclic order prime to } p"] \\
\{1\} \arrow[u, hook, "p\text{--group}"]
\end{tikzcd}$$

Corollary 3.9 *The group $G = Gal(K/\mathbf{Q}_p)$ is solvable. Moreover, its inertia subgroup G_0 has a normal Sylow p-subgroup (namely G_1) with cyclic quotient.*

To study continuous homomorphisms $\rho_p : G_{\mathbf{Q}_p} \to GL_2(R)$, we are looking at finite quotients $Gal(K/\mathbf{Q}_p)$ of $G_{\mathbf{Q}_p}$. If we can next define ramification subgroups for the infinite Galois group $G_{\mathbf{Q}_p}$, then we can describe properties of ρ_p in terms of its effect on these subgroups.

3.1 Infinite extensions

Let $\mathbf{Q}_p \subseteq M \subseteq L$ be finite Galois extensions, with respective valuation rings A_M, A_L. We have a restriction map

$$\begin{tikzcd}
Gal(L/\mathbf{Q}_p) \arrow[r] & Gal(M/\mathbf{Q}_p) \\
G_i^{(L)} \arrow[u, phantom, sloped, "\subseteq"] & G_i^{(M)}, \arrow[u, phantom, sloped, "\subseteq"]
\end{tikzcd}$$

where the horizontal map is restriction. In general the image of $G_i^{(L)}$ under this restriction map does not equal $G_i^{(M)}$; in order to have this happen, we need the upper numbering of ramification groups (see Chapter IV of [29]).

It is, however, true that the image of $G_i^{(L)}$ equals $G_i^{(M)}$ for $i = 0, 1$. This follows for $i = 0$ since $\sigma \in G_0^{(L)} \iff$ the valuation of $\sigma(x) - x > 0$ for all $x \in A_L$. It follows that the valuation of $\sigma|_M(x) - x > 0$ for all $x \in A_M$. The case $i = 1$ now follows since the Sylow p-subgroups of the image are the image of the Sylow p-subgroups.

Hence, we obtain, for $i = 0$ or 1, an inverse system consisting of the groups $G_i^{(L)}$ as L runs through all finite Galois extensions of $\mathbf{Q}_p$, together with homomorphisms $G_i^{(L)} \twoheadrightarrow G_i^{(M)}$ whenever $M \subseteq L$. Inside $\prod Gal(L/\mathbf{Q}_p)$, by definition of inverse limit, we have

$$\begin{array}{rcl}
G_{\mathbf{Q}_p} & = & \varprojlim Gal(L/\mathbf{Q}_p) \\
 & & \rotatebox{90}{$\subseteq$} \\
G_0 & = & \varprojlim G_0^{(L)} \\
 & & \rotatebox{90}{$\subseteq$}
\end{array}$$

$$G_1 \;=\; \varprojlim G_1^{(L)}$$

A second inverse system consists of the groups $Gal(L/\mathbf{Q}_p)/G_0^{(L)}$. Homomorphisms for $M \subseteq L$ are obtained from the restriction map and the fact that the image under restriction of $G_0^{(L)}$ is equal to $G_0^{(M)}$.

$$\begin{tikzcd}
Gal(L/\mathbf{Q}_p)/G_0^{(L)} \arrow[r] \arrow[d, equal] & Gal(M/\mathbf{Q}_p)/G_0^{(M)} \arrow[d, equal] \\
Gal(k_L/\mathbf{F}_p) \arrow[r] & Gal(k_M/\mathbf{F}_p)
\end{tikzcd}$$

We use this to show that

$$G_{\mathbf{Q}_p}/G_0 \cong Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p) \cong \hat{\mathbf{Z}} \cong \prod_q \mathbf{Z}_q.$$

We shall see later that

$$G_0/G_1 \cong \prod_{q \neq p} \mathbf{Z}_q.$$

The following will come in useful:

Theorem 3.10 *If H is a closed subgroup of $G_{\mathbf{Q}_p}$, then $H = Gal(\bar{\mathbf{Q}}_p/L)$, where L is the fixed field of H.*

Proof: Let H be a closed subgroup of $G_{\mathbf{Q}_p}$, and let L be its fixed field. Then for any finite Galois extension M/L, we obtain a commutative diagram

$$\begin{tikzcd}
H \arrow[r, hook] \arrow[dr] & Gal(\bar{\mathbf{Q}}_p/L) \arrow[d] \\
 & Gal(M/L)
\end{tikzcd}$$

Since $Gal(\bar{\mathbf{Q}}_p/L) = \varprojlim Gal(M/L)$, this implies that H is dense in $Gal(\bar{\mathbf{Q}}_p/L)$. Since H is closed, $H = Gal(\bar{\mathbf{Q}}_p/L)$. QED

In fact infinite Galois theory provides an order-reversing bijection between intermediate fields and the closed subgroups of the Galois group.

3.2 Structure of $G_{\mathbf{Q}_p}/G_0$ and $G_{\mathbf{Q}_p}/G_1$

We now have normal subgroups $G_1 \subseteq G_0 \subseteq G_{\mathbf{Q}_p}$. Suppose that $\rho : G_{\mathbf{Q}} \to GL_2(R)$ is one of the naturally occurring continuous homomorphisms in which we are interested, e.g. associated to an elliptic curve or modular form. We get continuous homomorphisms $\rho_p : G_{\mathbf{Q}_p} \to GL_2(R)$, such that $\rho_p(G_0) = \{1\}$ for all but finitely many p (in which case we say that ρ is *unramified at* p). For a p at which ρ is unramified, ρ induces a homomorphism $G_{\mathbf{Q}_p}/G_0 \to GL_2(R)$. Often it will be the case that $\rho_p(G_1) = \{1\}$ for a ramified p (in which case we say that ρ is *tamely ramified at* p). In this case ρ induces a homomorphism $G_{\mathbf{Q}_p}/G_1 \to GL_2(R)$. Thus, we are interested in the structure of the two groups $G_{\mathbf{Q}_p}/G_0$ and $G_{\mathbf{Q}_p}/G_1$.

Here is a rough outline of how we establish this. In the finite extension case considered already, G/G_0 is isomorphic to $Gal(k/\mathbf{F}_p)$ (and so is cyclic), and G_0/G_1 embeds in $k^\times$ (and so is cyclic of order prime to p). In the limit, $G_{\mathbf{Q}_p}/G_0 = \varprojlim G/G_0 = \varprojlim Gal(k/\mathbf{F}_p) = Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p)$ (and so is procyclic), and G_0/G_1

embeds in $\varprojlim k^\times$. A few things need to be checked along the way. The first statement amounts to there being one and only one homomorphism onto each $Gal(k/\mathbf{F}_p)$. For the second statement, we check that the restriction maps between the G_0/G_1 translate into the norm map between the $k^\times$. This gives an injective map $G_0/G_1 \to \varprojlim k^\times$, shown surjective by exhibiting explicit extensions of $\mathbf{Q}_p$.

As for $G_{\mathbf{Q}_p}/G_1$, consider the extension of groups at the finite level:

$$1 \to G_0/G_1 \to G/G_1 \to G/G_0 \to 1.$$

Thus, G/G_1 is *metacyclic*, i.e. has a cyclic normal subgroup with cyclic quotient. The group G/G_1 acts by conjugation on the normal subgroup G_0/G_1. Since G_0/G_1 is abelian, it acts trivially on itself, and thus the conjugation action factors through G/G_0. So G/G_0 acts on G_0/G_1, but since it is canonically isomorphic to $Gal(k/\mathbf{F}_p)$, it also acts on $k^\times$. These actions commute with the map $G_0/G_1 \to k^\times$. In the limit, $G_{\mathbf{Q}_p}/G_1$ is prometacyclic, and the extension of groups

$$1 \to G_0/G_1 \to G_{\mathbf{Q}_p}/G_1 \to G_{\mathbf{Q}_p}/G_0 \to 1$$

is a semidirect product since $G_{\mathbf{Q}_p}/G_0$ is free, with the action given by that of $Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p)$ on $\varprojlim k^\times$. In particular, the Frobenius map is a (topological) generator of $Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p)$ and acts by mapping elements to their pth power.

Theorem 3.11 *Let G_0, G_1 denote the 0th and 1st ramification subgroups of $G_{\mathbf{Q}_p}$. Then $G_{\mathbf{Q}_p}/G_0 \cong \varprojlim Gal(k/\mathbf{F}_p) \cong \hat{\mathbf{Z}}$ and $G_0/G_1 \cong \varprojlim k^\times \cong \prod_{q \neq p} Z_q$, where the maps in the inverse system are norm maps. Moreover, $G_{\mathbf{Q}_p}/G_1$ is (topologically) generated by two elements x, y where y generates G_0/G_1, x*

maps onto the Frobenius element in $G_{\mathbf{Q}_p}/G_0 = Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p)$, and $x^{-1}yx = y^p$.

For any finite extension $L/\mathbf{Q}_p$, define $G^{(L)} = Gal(L/\mathbf{Q}_p)$, A_L is the valuation ring of L, and k_L the residue field of L.

Let $M \subseteq L$ be finite Galois extensions of $\mathbf{Q}_p$. We obtain the following diagram:

$$\begin{tikzcd}
 & G_{\mathbf{Q}_p} \arrow[dl] \arrow[dr] \arrow[ddl, bend right, "\phi_L"'] \arrow[ddr, bend left, "\phi_M"] & \\
G^{(L)}/G_0^{(L)} \arrow[rr] \arrow[d, two heads] & & G^{(M)}/G_0^{(M)} \arrow[d, two heads] \\
Gal(k_L/\mathbf{F}_p) \arrow[rr] & & Gal(k_M/\mathbf{F}_p)
\end{tikzcd}$$

Note that G_0 lis in the kernel of each ϕ_L. From this diagram we get maps $\phi_L : G_{\mathbf{Q}_p}/G_0 \to Gal(k_L/\mathbf{F}_p)$. We note the following facts.

1) Given k, there is only one such map, ϕ_k. The reason here is that if $k_L = k_M = k$, then there is a unique map

$$Gal(LM/\mathbf{Q}_p) \to Gal(k_{LM}/\mathbf{F}_p) \to Gal(k/\mathbf{F}_p)$$

since $Gal(k_{LM}/\mathbf{F}_p)$ is cyclic.

Definition 3.12 *The valuation ring of the fixed field of ϕ_k will be denoted $W(k)$, the ring of infinite Witt vectors of k. An alternative explicit description is given in [17]. Note that $W(\mathbf{F}_p)$ is simply $\mathbf{Z}_p$.*

2) Given k, there is some L with $k_L = k$. (We may take $L = \mathbf{Q}_p(\zeta)$, where ζ is a primitive $(|k| - 1)$th root of 1.)

We consider the inverse system consisting of groups $Gal(k/\mathbf{F}_p)$, where $k/\mathbf{F}_p$ runs through finite Galois extensions, together with the usual restriction maps. Then $\{G_{\mathbf{Q}_p}/G_0, \{\phi_k\})$ is an object of the new category associated to this inverse system. We obtain a homomorphism

$$G_{\mathbf{Q}_p}/G_0 \to \varprojlim Gal(k/\mathbf{F}_p) \cong \hat{\mathbf{Z}},$$

and this is an isomorphism because of (1) and (2) above.

Next we study the structure of G_0/G_1. We have the following diagram:

$$\begin{tikzcd}
 & G_0 \arrow[dl] \arrow[dr] \arrow[ddl, bend right, "\psi_L"'] \arrow[ddr, bend left, "\psi_M"] & \\
G_0^{(L)}/G_1^{(L)} \arrow[rr] \arrow[d] & & G_0^{(M)}/G_1^{(M)} \arrow[d] \\
k_L^\times \arrow[rr] & & k_M^\times
\end{tikzcd}$$

Note that G_1 lies in the kernel of each ψ_L. It is a simple exercise to show that the norm map $k_L^\times \to k_M^\times$ makes the bottom square commutative.

Theorem 3.13 *There is a canonical isomorphism* $G_0/G_1 \cong \varprojlim k^\times$.

Proof: Let $L/\mathbf{Q}_p$ be a finite Galois extension, and recall that

$$G_0^{(L)}/G_1^{(L)} \hookrightarrow k_L^\times$$

naturally. Then

$$G_0 \to G_0^{(L)} \to k_L^\times$$

factors through G_0/G_1. In this way we get maps $G_0/G_1 \to k_L^\times$ for each L, and finally a map $G_0/G_1 \to \varprojlim k^\times$, where the inverse limit is taken over all finite Galois extensions $k/\mathbf{F}_p$, and for $k_1 \subseteq k_2$, the map $k_2^\times \to k_1^\times$ is the norm map. The fact that

$G_0/G_1 \to \varprojlim k^\times$ is an isomorphism, follows by exhibiting fields $L/\mathbf{Q}_p$ with $G_0^{(L)}/G_1^{(L)} \cong k^\times$ for any finite field $k/\mathbf{F}_p$, namely $L = Q_p(\zeta, p^{1/d})$ where $d = |k| - 1$ and ζ is a primitive dth root of 1 . QED

Note: the fields exhibited above yield (surjective) *fundamental characters* $G_0/G_1 \to k^\times$. Let μ_n be the group of nth roots of 1 in $\bar{\mathbf{F}}_p$, where $(p, n) = 1$. For $m|n$, we have a group homomorphism $\mu_n \to \mu_m$ given by $x \mapsto x^{n/m}$, which forms an inverse system. Let $\mu = \varprojlim \mu_n$.

Lemma 3.14 *We have a canonical isomorphism $\varprojlim k^\times \cong \mu$.*

Proof: Since the groups $k^\times$ form a subset of the groups μ_n, we obtain a surjection from $\varprojlim \mu_n \to \varprojlim k^\times$. To obtain a map in the other direction, note that the numbers $p^f - 1$ are cofinal in the set of integers prime to p. In fact, if d is such an integer integer, there is some $f \geq 1$ with $p^f \equiv 1 \pmod{d}$, e.g. $f = \varphi(d)$. QED

μ is noncanonically isomorphic to $\prod_{q \neq p} \mathbf{Z}_q$, since μ_n is noncanonically isomorphic to $\mathbf{Z}/n\mathbf{Z}$, and

$$\varprojlim \mathbf{Z}/n\mathbf{Z} \cong \prod_{q \neq p} \mathbf{Z}_q,$$

since $\mathbf{Z}/n\mathbf{Z} \cong \prod \mathbf{Z}/p_i^{r_i}\mathbf{Z}$ for $n = \prod p_i^{r_i}$.

We now examine the map

$$G_0^{(L)}/G_1^{(L)} \hookrightarrow k_L^\times.$$

One can check that it is $G^{(L)}$-equivariant (with the conjugation action on the left, and the natural action on the right). Now $G_0^{(L)}$ acts trivially on the left ($G_0^{(L)}/G_1^{(L)}$ is abelian), and also trivially on the right by definition. Hence we end up with a

$G^{(L)}/G_0^{(L)}(\cong Gal(k_L/\mathbf{F}_p))$-equivariant map. The canonical isomorphism $G_0/G_1 \to \varprojlim k^\times$ is $G_{\mathbf{Q}_p}/G_0$-equivariant.

The upshot is that $G_{\mathbf{Q}_p}/G_1$ is topologically generated by 2 elements x, and y, where x generates a copy of $\hat{\mathbf{Z}}$, y generates a copy of $\prod_{q\neq p}\mathbf{Z}_q$, with one relation $x^{-1}yx = y^p$. This is seen from the short exact sequence of groups:

$$1 \to G_0/G_1 \to G_{\mathbf{Q}_p}/G_1 \to G_{\mathbf{Q}_p}/G_0 \to 1,$$

in which $G_0/G_1 \cong \prod_{q\neq p}\mathbf{Z}_q$, and $G_{\mathbf{Q}_p}/G_0 \cong \hat{\mathbf{Z}}$ are both procyclic (i.e topologically generated by one element). Since $\hat{\mathbf{Z}}$ is free, the sequence is split, i.e. defines a semidirect product with the action given as above.

Next, let $\rho : G_{\mathbf{Q}} \to GL_n(k)$ be given, where k is a finite field of characteristic ℓ. The image of ρ is finite, say $Gal(K/\mathbf{Q})$, where K is a number field. Letting the finite set of rational primes ramified in $K/\mathbf{Q}$ be S, we see that ρ is unramified at every $p \notin S$, i.e. $\rho_p(G_0) = \{1\}$ and so ρ_p factors through $G_{\mathbf{Q}_p}/G_0$ for all $p \notin S$. Consider next $\rho_\ell : G_{\mathbf{Q}_\ell} \to GL_n(k)$.

Call ρ_ℓ *semisimple* if $V := k^n$, viewed as a $k[G_{\mathbf{Q}_\ell}]$-module, is semisimple, i.e. a direct sum of irreducible modules.

Theorem 3.15 *If ρ_ℓ is semisimple, then $\rho_\ell(G_1) = \{1\}$, and so ρ_ℓ factors through $G_{\mathbf{Q}_\ell}/G_1$.*

Proof: Assume V is irreducible, i.e. has no proper $k[G_{\mathbf{Q}_\ell}]$-submodules. Let $V' = \{v \in V | g(v) = v \text{for all} g \in \rho_\ell(G_1)\}$. Since G_1 is a pro-ℓ group, $\rho_\ell(G_1)$ is a finite ℓ-group and so its orbits on V are of length 1 or a power of ℓ. The orbits of length 1 comprise V', so that $|V'| \equiv |V| \pmod{\ell} \equiv 0 \pmod{\ell}$. In particular, $V' \neq \{0\}$. Since G_1 is normal in $G_{\mathbf{Q}_\ell}$, V' is stable under $G_{\mathbf{Q}_\ell}$, implying by irreducibility of V that $V' = V$. Thus $\rho_\ell(G_1)$ acts trivially on the whole of V. For a semisimple V, we apply the above to each summand of V. QED

The Big Picture. We have defined subgroups $G_{\mathbf{Q}_p}$ (one for each prime p) of $G_{\mathbf{Q}}$ with much simpler structure than $G_{\mathbf{Q}}$ itself. This will allow us to describe representations $\rho : G_{\mathbf{Q}} \to GL_n(R)$ in terms of their restrictions ρ_p to these subgroups. Each ρ_p can be described in turn by its effect on ramification subgroups of $G_{\mathbf{Q}_p}$, ultimately enabling us to define useful numerical invariants associated to ρ (see chapter 5). First, however, we need some natural sources of Galois representations ρ and these will be provided by elliptic curves, modular forms, and more generally group schemes. Φ

4

Galois Representations from Elliptic Curves, Modular Forms, and Group Schemes

Having introduced Galois representations, we next describe natural sources for them, that will lead to the link with Fermat's Last Theorem.

4.1 Elliptic curves

An *elliptic curve* over $\mathbf{Q}$ is given by equation $y^2 = f(x)$, where $f \in \mathbf{Z}[x]$ is a cubic polynomial with no repeated roots in $\bar{\mathbf{Q}}$. A good reference for this theory is [32].

Example: $y^2 = x(x-1)(x+1)$

If K is a field, set $E(K) := \{(x,y) \in K \times K | y^2 = f(x)\} \cup \{\infty\}$. We begin by studying $E(\mathbf{C})$.

Let Λ be a lattice inside $\mathbf{C}$. Define the *Weierstrass $\wp$-function* by

$$\wp(z;\Lambda) = \frac{1}{z^2} + \sum_{0 \neq \omega \in \Lambda} (\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2})$$

and the $2k$th *Eisenstein series* by

$$G_{2k}(\Lambda) = \sum_{0\neq\omega\in\Lambda} \omega^{-2k}.$$

These arise as the coefficients in the Laurent series of $\wp$ about $z = 0$, namely:

$$\wp(z) = \frac{1}{z^2} + 3G_4 z^2 + 5G_6 z^4 + 7G_8 z^6 + \dots,$$

which is established by rearranging the infinite sum, allowed by the following.

Proposition 4.1 *The function $\wp$ is absolutely and locally uniformly convergent on $\mathbf{C} - \Lambda$. It has poles exactly at the points of Λ and all the residues are 0. $G_{2k}(\Lambda)$ is absolutely convergent if $k > 1$.*

Definition 4.2 *A function f is called elliptic with respect to Λ (or doubly periodic) if $f(z) = f(z+\omega)$ for all $\omega \in \Lambda$. Note that to check ellipticity it suffices to show this for $\omega = \omega_1, \omega_2$, the fundamental periods of Λ. An elliptic function can be regarded as a function on $\mathbf{C}/\Lambda$.*

Proposition 4.3 *$\wp$ is an even function and elliptic with respect to Λ.*

Proof: Clearly, $\wp$ is even, i.e. $\wp(z) = \wp(-z)$. By local uniform convergence, we can differentiate term-by-term to get $\wp'(z) = -2\sum_{\omega\in\Lambda} \frac{1}{(z-\omega)^3}$, and so $\wp'$ is elliptic with respect to Λ. Integrating, for each $\omega \in \Lambda$, $\wp(z+\omega) = \wp(z) + C(\omega)$, where $C(\omega)$ does not depend on z. Setting $z = -\omega_i/2$ and $\omega = \omega_i$ ($i = 1$ or 2) gives

$$\wp(\omega_i/2) = \wp(-\omega_i/2) + C(\omega_i) = \wp(\omega_i/2) + C(\omega_i),$$

using the evenness of $\wp$. Thus, $C(\omega_i) = 0 (i = 1, 2)$, so $\wp$ is elliptic. QED

Let $g_2 = 60G_4(\Lambda)$ and $g_3 = 140G_6(\Lambda)$.

Proposition 4.4 $(\wp'(z))^2 = 4\wp(z)^3 - g_2\wp(z) - g_3 (z \notin \Lambda).(*)$

Proof: Let $f(z) = (\wp'(z))^2 - (a\wp(z)^3 - b\wp(z)^2 - c\wp(z) - d)$. Considering its explicit Laurent expansion around $z = 0$, since f is even, there are terms in $z^{-6}, z^{-4}, z^{-2}, z^0$ whose coefficients are linear in a, b, c, d. Choosing $a = 4, b = 0, c = g_2, d = g_3$ makes these coefficients zero, and so f is holomorphic at $z = 0$ and even vanishes there. We already knew that f is holomorphic at points not in Λ. Since f is elliptic, it is bounded. By Liouville's theorem, bounded holomorphic functions are constant. $f(0) = 0$ implies this constant is 0. QED

Theorem 4.5 *The discriminant* $\Delta(\Lambda) := g_2^3 - 27g_3^2 \neq 0$, *and so* $4x^3 - g_2x - g_3$ *has distinct roots, whence* E_Λ *defined by* $y^2 = 4x^3 - g_2x - g_3$ *is an elliptic curve (over* $\mathbf{Q}$ *if* $g_2, g_3 \in \mathbf{Q}$*).*

Proof: Setting $\omega_3 = \omega_1 + \omega_2$, $\wp'(\omega_i/2) = -\wp'(-\omega_i/2) = -\wp'(\omega_i/2)$ since $\wp'$ is odd and elliptic. Hence $\wp'(\omega_i/2) = 0$. Thus $4\wp(\omega_i/2)^3 - g_2\wp(\omega_i/2) - g_3 = 0$. We just need to show that these three roots of $4x^3 - g_2x - g_3 = 0$ are distinct.

Consider $\wp(z) - \wp(\omega_i/2)$. This has exactly one pole, of order 2, and so by the next lemma has either two zeros of order 1 (which cannot be the case since this function is even and elliptic) or 1 zero of order 2, namely $\omega_i/2$. Hence $\wp(\omega_i/2) - \wp(\omega_j/2) \neq 0$ if $i \neq j$. QED

Lemma 4.6 *If* f *is elliptic and* $v_w(f)$ *is the order of vanishing of* f *at* w, *then* $\sum_{w \in \mathbf{C}/\Lambda} v_w(f) = 0$. *Moreover, the sum of all the zeros minus the poles is* $0 \pmod{\Lambda}$*.*

Proof: By Cauchy's residue theorem, $\sum_{w\in\mathbf{C}/\Lambda} v_w(f) = \frac{1}{2\pi i}\int \frac{f'}{f}dz$, where the integral is over the boundary of the fundamental parallelogram with vertices $0, \omega_1, \omega_2, \omega_3$ (if a pole or zero happens to land on the boundary, then translate the whole parallelogram to avoid it). By ellipticity, the contributions from parallel sides cancel, so the integral is 0. The last statement is proved similarly, using $\frac{1}{2\pi i}\int z\frac{f'}{f}dz$. QED

Theorem 4.7 *There is a bijection (in fact a homeomorphism of Riemann surfaces) $\phi : \mathbf{C}/\Lambda \to E_\Lambda(\mathbf{C})$ given by*

$$z \mapsto (\wp(z), \wp'(z))(z \notin \Lambda), z \mapsto \infty(z \in \Lambda).$$

Proof: Ellipticity of $\wp$ and $\wp'$ implies that ϕ is well-defined and $(*)$ shows that the image is in $E_\Lambda(\mathbf{C})$. To show surjectivity, given $(x, y) \in E_\Lambda(\mathbf{C}) - \{\infty\}$, we consider $\wp(z) - x$, a nonconstant elliptic function with a pole (at 0) and so a zero, say at $z = a$. By $(*)$, $\wp'(a)^2 = y^2$. By oddness of $\wp'$ and evenness of $\wp$, we see that $\phi(a)$ or $\phi(-a)$ is (x, y).

To show injectivity, if $\phi(z_1) = \phi(z_2)$ with $2z_1 \notin \Lambda$, then consider $\wp(z) - \wp(z_1)$, which has a pole of order 2 and zeros at $z_1, -z_1, z_2$, so $z_2 \equiv \pm z_1 \pmod{\Lambda}$. If also $\phi'(z_1) = \phi'(z_2)$, then this fixes the sign. QED

Note that this bijection from $\mathbf{C}/\Lambda$, which is a group (in fact a torus), puts a group structure on $E_\Lambda(\mathbf{C})$, so that it is isomorphic to $\mathbf{R}/\mathbf{Z} \times \mathbf{R}/\mathbf{Z}$. Our desired Galois representations will come from Galois actions on certain finite subgroups of $E_\Lambda(\mathbf{C})$. Later, we shall see that every elliptic curve over $\mathbf{C}$ is of the form E_Λ.

Theorem 4.8 *The group law on $E_\Lambda(\mathbf{C})$ is given by saying that three points P_1, P_2, P_3 add up to the identity, ∞, if and only if P_1, P_2, P_3 are collinear. If two of the points coincide, this means* *the tangent at that point.*

Proof: Fixing z_1, z_2, let $y = mx + b$ be the line through $P_1 = (\wp(z_1), \wp'(z_1))$ and $P_2 = (\wp(z_2), \wp'(z_2))$.

Consider $f(z) = \wp'(z) - m\wp(z) - b$, which has one pole of order 3 at 0, whence three zeros. Two of these are z_1, z_2 - let the third be z_3. Since the zeros minus poles equals 0, we get $z_1 + z_2 + z_3 = 0$. Thus, if $P_3 = (\wp(z_3), \wp'(z_3)$, then $P_1 + P_2 + P_3 = \infty$. QED

The importance of this is that it means that the coordinates of $(x_1, y_1) + (x_2, y_2)$ are rational functions in $x_1, x_2, y_1, y_2, g_2, g_3$. This yields a group structure on $E(K)$ whenever $g_2, g_3 \in K$, since e.g. the associative law is a formal identity in these rational functions.

Definition 4.9 *Given an elliptic curve E over $\mathbf{Q}$ and positive integer n, the n-division points of E are given by $E[n] := \{P \in E(\mathbf{C}) | nP = \infty\}$.*

Example: If E is $y^2 = f(x) \in \mathbf{Z}[x]$, let α_i $(i = 1, 2, 3)$ be the roots of $f(x) = 0$ and $P_i = (\alpha_i, 0)$. The tangent at P_i is vertical and so P_i, P_i, ∞ are collinear, whence $2P_i = \infty$. Thus $E[2] = \{\infty, P_1, P_2, P_3\}$. (No other point has order 2 by the following.)

Lemma 4.10 *Let E be an elliptic curve over $\mathbf{Q}$. Since $\mathbf{C}/\Lambda \cong \mathbf{R}/\mathbf{Z} \times \mathbf{R}/\mathbf{Z}$, $E[n] \cong \mathbf{Z}/n \times \mathbf{Z}/n$. Moreover, there are polynomials $f_n \in \mathbf{Q}[x]$ such that $E[n] = \{(x, y) | f_n(x) = 0\} \cup \{\infty\}$.*

Proof: The elements of $\mathbf{R}/\mathbf{Z}$ of order dividing n form a cyclic group of order n. The f_n come from iterating the rational function that describes addition of two points on the curve. QED

Example: Continuing the case $n = 2$, we see that $f_2 = f$.

By the last lemma, $E[n] \subseteq E(\bar{\mathbf{Q}})$ and if $P = (x, y) \in E[n]$ and $\sigma \in G_{\mathbf{Q}}$, then $\sigma(P) = (\sigma(x), \sigma(y) \in E[n]$. This action of $G_{\mathbf{Q}}$ on $E[n] \cong \mathbf{Z}/n \times \mathbf{Z}/n$ produces a homomorphism $\rho_{E,n} : G_{\mathbf{Q}} \to GL_2(\mathbf{Z}/n)$. These are the Galois representations associated to E.

Let ℓ be a prime. Consider the inverse system consisting of groups $GL_2(\mathbf{Z}/\ell^n)$ together with the natural maps between them. An object of the corresponding new category is given by $(G_{\mathbf{Q}}, \{\rho_{E,\ell^n}\})$, yielding a homomorphism $\rho_{E,\ell^\infty} : G_{\mathbf{Q}} \to GL_2(\mathbf{Z}_\ell)$, called the *$\ell$-adic representation associated to E*. Another way of viewing this is as the Galois action on the *ℓ-adic Tate module* $T_\ell(E) = \varprojlim E[\ell^n]$, where the maps in the inverse system are $E[\ell^n] \to E[\ell^m]$ for $n > m$ defined by $P \mapsto \ell^{n-m}P$.

Example: Continuing the case $n = 2$, note that $GL_2(\mathbf{Z}/2) \cong S_3$, the symmetric group on 3 letters. The action of $G_{\mathbf{Q}}$ on $E[2] = \{\infty, (\alpha_1, 0), (\alpha_2, 0), (\alpha_3, 0)\}$ amounts to permuting the roots α_i of f, and so the image of $\rho_{E,2}$ is $Gal(K/\mathbf{Q}) \leq S_3$ where K is the splitting field of f.

Given a typical E (meaning one with no complex multiplications), ρ_{E,ℓ^∞} is surjective for all but finitely many primes ℓ. In fact it is surjective for all ℓ for a set of elliptic curves of density 1, for example if E is the curve $y^2 + y = x^3 - x$ (elliptic - you can complete the square).

These ℓ-adic Galois representations encode much information about elliptic curves. For example, for a fixed ℓ, ρ_{E,ℓ^∞} and ρ_{E',ℓ^∞} are equivalent (conjugate) if and only if E and E' are isogenous. Later on, we shall study in great detail the ℓ-adic representations associated to semistable elliptic curves.

4.2 Group schemes

An elliptic curve E defined over $\mathbf{Q}$ yields groups $E(A)$ for any $\mathbf{Q}$-algebra A. This can be usefully generalized as follows. Ultimately we shall define finite, flat group schemes and see that they provide a quite general source of Galois representations. A good resource for this section is [37].

Definition 4.11 *Let R be a commutative ring with 1. An affine group scheme over R is a representable functor from the category of R-algebras (i.e. rings A together with a homomorphism $R \to A$ with morphisms ring homomorphisms that make a commutative triangle over R) to the category of groups.*

Recall that a functor F is *representable* by the R-algebra $\Re$ if and only if $F(A) = \hom_{R-alg}(\Re, A)$. In general, a representable functor from the category of R-algebras to the category of sets actually defines an affine scheme over R - the added group structure gives the representing ring $\Re$ the structure of a Hopf algebra.

Example: (i) The functor $\mathbf{G}_a$ given by $\mathbf{G}_a(A) = A^+$ is representable since $A^+ = \hom_{R-alg}(R[T], A)$ (an R-algebra homomorphism from $R[T]$ to A is determined by whatever T maps to, and this can be any element of A).

(ii) The functor $\mathbf{G}_m$ given by $\mathbf{G}_m(A) = A^\times$(the units of A) is representable since $A^\times = \hom_{R-alg}(R[X,Y]/(XY-1), A)$ (under an R-algebra homomorphism, X has to map to something invertible in A and then Y maps to its inverse).

(iii) More generally, the functor GL_n is representable since $GL_n(A) = \hom_{R-alg}(R[T_{11}, T_{12}, ..., T_{nn}, Y]/(det((T_{ij}))Y-1), A)$. Note $GL_1 = \mathbf{G}_m$.

(iv) The functor μ_n defined by $\mu_n(A) := \{x \in A | x^n = 1\}$ is representable since $\mu_n(A) = \hom_{R-alg}(R[T]/(T^n - 1), A)$. μ_n is a subgroup scheme of $\mathbf{G}_m$.

The following example shows how certain group schemes can give rise to Galois representations. It will be generalized below.

Definition 4.12 *Let K be a field, $\ell \neq \mathrm{char} K$. Then G_K acts on $\mu_{\ell^n}(\bar{K}) \cong \mathbf{Z}/\ell^n$. This yields a representation $\chi_{\ell^n} : G_K \to GL_1(\mathbf{Z}/\ell^n) = (\mathbf{Z}/\ell^n)^\times$. Putting these together yields $\chi_{\ell^\infty} : G_K \to GL_1(\mathbf{Z}_\ell) = \mathbf{Z}_\ell^\times$, called the ℓ-adic cyclotomic character. As with elliptic curves, we can provide an alternative defintion by doing the inverse limit before the Galois action; namely letting $T_\ell(\mu) := \varprojlim \mu_{\ell^n}$, then χ_{ℓ^∞} gives the action of G_K on the Tate module $T_\ell(\mu)$.*

Exercise: Let $K = \mathbf{Q}$. Show that the ℓ-adic cyclotomic character χ is unramified at all primes $p \neq \ell$, i.e. $\chi_p : G_{\mathbf{Q}_p} \to \mathbf{Z}_\ell^\times$ factors through the inertia subgroup G_0. This then induces a map $Gal(\bar{\mathbf{F}}_p/\mathbf{F}_p) \cong G_{\mathbf{Q}_p}/G_0 \to \mathbf{Z}_\ell^\times$. Show that the image of the pth Frobenius element is p. (Hint: note that $\sigma(\zeta) = \zeta^{\chi(\sigma)}$ for any ℓ-power root of 1 in $\bar{\mathbf{Q}}$.)

Exercise: Some universal ring constructions.

(i) Given ring homomorphism $R \to S$ and R-algebra A, consider the collection of rings B and homomorphisms that make the following diagram commute:

$$\begin{tikzcd}
A \arrow[r] & B \\
R \arrow[u] \arrow[r] & S \arrow[u]
\end{tikzcd}$$

Show that these form the objects in a category of S-algebras, with an initial object. This initial object is the *tensor product* $A \otimes_R S$. In particular, $R[T_1, \ldots, T_n] \otimes_R S = S[T_1, \ldots, T_n]$.

(ii) Given $x \in R$, consider the collection of R-algebras A such that under $R \to A$, x maps to a unit in A. Show that these form the objects in a category, with an initial object. This initial object is the *localization* of R at x.

The case of (i) we shall be most interested in is where $R = \mathbf{Z}_\ell$. If $A \cong \mathbf{Z}_\ell^n \times T$ (T torsion) as a $\mathbf{Z}_\ell$-module, then $A \otimes_R \mathbf{Q}_\ell \cong \mathbf{Q}_\ell^n$ as a $\mathbf{Q}_\ell$-module and $A \otimes_R \mathbf{F}_\ell \cong \mathbf{F}_\ell^n \times T/\ell T$ as an $\mathbf{F}_\ell$-module.

Let $\phi : R \to S$ be a ring homomorphism. Given a functor F on R-algebras, we get a functor F' on S-algebras since via composition with ϕ every S-algebra $S \to A$ is an R-algebra. If $\Re$ is an R-algebra, then $\hom_{S-alg}(\Re \otimes_R S, A) \cong \hom_{R-alg}(\Re, A)$ (adjoint associativity - follows quickly from universal description of tensor above), so if F is representable (by R-algebra $\Re$), then F' is representable (by S-algebra $\Re \otimes_R S$). This process is called *base change*.

A *homomorphism of affine group schemes* over R is a natural map $G \to H$. This gives us for each R-algebra A a group homomorphism $G(A) \to H(A)$ such that whenever $A \to B$ is a homomorphism of R-algebras, the following diagram commutes:

$$\begin{tikzcd}
G(A) \arrow[r] \arrow[d] & H(A) \arrow[d] \\
G(B) \arrow[r] & H(B)
\end{tikzcd}$$

Here the vertical maps come from G and H being functors.

Exercise: For each R-algebra A, define $F(A) = \ker(G(A) \to H(A))$. Show that F is a group scheme over R.

For example, the determinant map gives a homomorphism from GL_n to $\mathbf{G}_m$ with kernel SL_n, and the map $x \mapsto x^n$ a homomorphism from $\mathbf{G}_m$ to $\mathbf{G}_m$ with kernel μ_n. It requires a lot more work to give the cokernel a functorial description.

We now have the objects and morphisms of the category of affine group schemes over a fixed ring R. An elliptic curve $E : y^2 = f(x)$ over $\mathbf{Q}$ gives for each $\mathbf{Q}$-algebra A a group $E(A)$ and for each $\mathbf{Q}$-algebra map $A \to B$ a group homomorphism $E(A) \to E(B)$. Is it an affine group scheme? Actually not - the obvious try is $\Re = \mathbf{Q}[x, y]/(y^2 - f(x))$, in which case $\hom_{\mathbf{Q}-alg}(\Re, A)$ yields the points (x, y) with coordinates in A satisfying $y^2 = f(x)$. This, however, misses the point at ∞. In other words, it shows that $E(A) - \{\infty\}$ defines an affine scheme. The answer is to define group schemes in general, obtained by patching together affine schemes called *charts.* For instance, $E(A) - \{\infty\}$ provides such a chart. We say more about non-affine group schemes later.

It turns out that the n-division points are nicer, since they define affine group schemes.

Definition 4.13 *An R-algebra A is called finite if it is finitely generated as an R-module. (Note that this is stronger than being finitely generated as an R-algebra - consider e.g. $R[T]$.) Let G be an affine group scheme over R. We call G finite over R if its representing ring $\Re$ is a finite R-algebra.*

In fact, any group scheme finite over R is affine. In particular, if $E : y^2 = f(x)$ is an elliptic curve over $\mathbf{Q}$ and ℓ a prime such that $f(x) \pmod{\ell}$ has distinct roots, then $E[n]$ is a finite (of rank n^2) affine group scheme over $\mathbf{Q}_p$. (It can be shown directly to be affine by finding a polynomial f whose zero set $V((f))$ - see below - does not meet $E[n]$, whence $E[n]$ lies in an affine chart U_f. See Conrad's article in [5] for details. A nice explicit description of $E[n]$ can be found in [9].) This will be discussed in the section on reduction of elliptic curves in the next chapter.

Next, we make explicit the kind of group schemes that yield Galois representations useful to us.

Definition 4.14 *Let G be a finite group scheme over a field K, e.g. represented by K-algebra $\Re$. We call G étale if $\Re$ is an étale (or separable) K-algebra, i.e. $\Re \otimes_K \bar{K} \cong \bar{K} \times \ldots \times \bar{K}$.*

Example: Let $G = \mu_\ell$, ℓ prime. We saw that $\Re = K[T]/(T^\ell - 1)$. Then $\Re \otimes_K \bar{K} = \bar{K}[T]/(T^\ell - 1)$. If char $K = \ell$, then this is $\bar{K}[T]/(T-1)^\ell$, which contains nilpotents and so G is not étale. If char $K \neq \ell$, then by Chinese remainder this is $\prod_{i=0}^{\ell-1} \bar{K}[T]/(T - \zeta^i) \cong \bar{K} \times \ldots \times \bar{K}$, and so G is étale.

Note that the étale situation here is that in which we obtained useful Galois representations earlier, namely the cyclotomic characters. This is clarified as follows (more details can be found in [37]).

Theorem 4.15 *The category of étale group schemes over K is equivalent to the category of finite groups with G_K acting continuously.*

Proof: Given an étale group scheme G, represented say by $\Re$, an étale K-algebra, consider the group $G(\bar{K}) = \hom_{K-alg}(\Re, \bar{K})$. Since $\Re$ is étale, $|\hom_{K-alg}(\Re, \bar{K})| = \dim_K \Re$, so $G(\bar{K})$ is a finite group. Given a K-algebra $\Re \to \bar{K}$ and $\sigma \in G_K$, the action is given by composing to get $\Re \to \bar{K} \xrightarrow{\sigma} \bar{K}$. The images of $\Re$ all lie in some finite Galois extension of K and so the action is continuous.

Conversely, given a finite group H with continuous G_K-action, consider first the case of transitive action, say $H = G_K h$. By continuity, choose finite extension L of K such that the action of G_K factors through $Gal(L/K)$. Let S be the subgroup fixing h and $\Re \subseteq L$ its fixed field. By Galois theory, all maps $\Re \to \bar{K}$ map to L and are conjugate, yielding a G_K-isomorphism $H \to \hom_{K-alg}(\Re, \bar{K})$ by sending h to one of them. If the action is intransitive, we obtain for each orbit a ring $\Re_i$ and then $\prod \Re_i$ works.

QED

Definition 4.16 *An R-algebra A is called flat if tensoring with A is an exact functor, i.e. if for any short exact sequence $0 \to M \to N \to L \to 0$ of R-modules, $0 \to M \otimes_R A \to N \otimes_R A \to L \otimes_R A \to 0$ is also exact. A finite group scheme over* Spec$R$ *is called flat if its representing ring A is a flat R-algebra.*

If $R = \mathbf{Z}_\ell$, then finite flat is equivalent to A being free of finite rank over R.

Let $\rho : G_{\mathbf{Q}_\ell} \to GL_n(\mathbf{Z}/\ell^m)$. The above theorem associates to ρ an étale group scheme G over $\mathbf{Q}_\ell$. Call ρ *good* if there is a finite flat group scheme F over $\mathbf{Z}_\ell$ such that the base change of F under $\mathbf{Z}_\ell \to \mathbf{Q}_\ell$ is G. This will be needed in defining when a Galois representation is semistable at ℓ in the next chapter.

4.3 General schemes

We next reconcile the above with the usual approach to schemes. We begin by defining affine schemes.

Let R be a commutative ring with 1 and Spec$R$ denote the set of prime ideals of R. For example, Spec$\mathbf{Z}_p = \{(0), p\mathbf{Z}_p\}$. Then Spec$R$ comes with a topology, the *Zariski topology*, defined by having the closed sets be the sets $V(I)$ as I runs through all ideals of R, where

$$V(I) := \{\wp \in \mathrm{Spec}R | I \subseteq \wp\}.$$

Exercise: Show that this does indeed define a topology on Spec$R$, i.e. that $\emptyset, R$ are closed sets and that arbitrary intersections and finite unions of closed sets are closed. Show that Spec$\mathbf{Z}_p$ is not Hausdorff. Show that Spec$R$ is always compact.

If $f : R \to S$ is a ring homomorphism and $\wp$ is a prime ideal of S, then $R/f^{-1}(\wp) \to S/\wp$ is an injective homomorphism into an integral domain, and so $f^{-1}(\wp)$ is a prime ideal of R. Thus f induces a map Spec$S \to$ Spec$R$, which can be checked to be continuous with respect to the Zariski topologies. For example, if I is an ideal, then since the prime ideals of R containing I are in bijection with those of R/I, Spec$(R/I) \to$ Spec$R$ is an injection with image $V(I)$. If $x \in R$, then likewise Spec$R_x \to$ Spec$R$ (from localization) is injective with image Spec$R - V((x))$. We thus think of Spec$R$ as a *ringed space* by having R_x be the ring of functions on this basic open set.

We shall identify Spec$\mathfrak{R}$ with the affine scheme represented by $\mathfrak{R}$. This is fair since to each R-algebra A, Spec$\mathfrak{R}$ associates the set Spec$(A \otimes_R \mathfrak{R})$, closely related to (but not exactly the same as) $\hom_{R-alg}(\mathfrak{R}, A)$.

Note that if A is an R-algebra, then the map $R \to A$ induces a map Spec$A \to$ Spec$R$, and Spec$A$ will be called an affine scheme over Spec$R$ via this map. Base change to Spec$\mathfrak{R}$ replaces Spec$A$ by Spec$(A \otimes_R \mathfrak{R})$. An R-algebra homomorphism $A \to B$

yields a commutative diagram:

$$\begin{tikzcd}
\mathrm{Spec}B \arrow[rr, "\phi"] \arrow[dr, "\phi_i"'] & & \mathrm{Spec}A \arrow[dl, "\pi_i"] \\
& \mathrm{Spec}R &
\end{tikzcd}$$

i.e. a morphism of affine schemes over Spec$R$. The category of affine schemes over Spec$R$ is hereby anti-equivalent to the category of R-algebras.

Exercise: Let $\wp \in \mathrm{Spec}R$ and let $\kappa(\wp) = \mathrm{Frac}(R/\wp)$. Show that $\kappa(\wp)$ is an R-algebra, so that Spec$\kappa(\wp)$ embeds in Spec$R$. Let A be an R-algebra, so that Spec$A$ is a cover of Spec$R$. Show that its fibre over $\wp$ can be identified with Spec$(A \otimes_R \kappa(\wp)$.

A *scheme* then is a ringed space admitting a covering by open sets that are affine schemes. Morphisms of schemes are defined locally, i.e. $f : S' \to S$ is a morphism if there is a covering of S by open, affine subsets SpecR_i such that $f^{-1}(\mathrm{Spec}R_i)$ is an affine scheme SpecR'_i and the restriction map Spec$R'_i \to$ SpecR_i is a morphism of affine schemes. We say that S' is a *scheme over* S. In Grothendieck's approach, this relative notion is important rather than absolute questions about a scheme. Questions about f turn into questions about the ring maps $f_i : R_i \to R'_i$. In particular, we say that f has property $(*)$ (for example, is finite or flat), if there is a covering of S such that each of the ring maps f_i has this property.

If S is a scheme, then a *group scheme over* S is a representable functor F from the category of schemes over S to the category of groups, i.e. there exists some scheme $\mathcal{S}$ over S such that $F(X) = \hom_{S-schemes}(X, S)$. For example, an elliptic curve over $\mathbf{Q}$ is a (non-affine) group scheme over Spec$\mathbf{Q}$.

4.4 Modular forms

Another source of Galois representations (in fact, which turns out to produce *all* that we are interested in) is modular forms. For this section, [28], [19], and [11] are recommended. Note that elliptic curves will ultimately correspond to modular forms of weight 2 and trivial Nebentypus, and so those forms will be highlighted.

Fix positive integers k, N and homomorphism $\epsilon : (\mathbf{Z}/N)^\times \to \mathbf{C}^\times$. Let $\mathcal{H} = \{z \in \mathbf{C} | \mathrm{Im}(z) > 0\}$. For $\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{Z})$, set $\sigma z = \frac{az+b}{cz+d}$ and $f|[\sigma]_k(z) = (cz+d)^{-k} f(\sigma z)$.

Set

$$\Gamma_0(N) := \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{Z}) | c \equiv 0 \pmod{N}\}$$

and

$$\Gamma_1(N) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) | d \equiv 1 \pmod{N}\}.$$

Definition 4.17 *A modular function of weight k, level N, and Nebentypus ϵ, is*

(i) a meromorphic function on $\mathcal{H}$,

(ii) satisfies $f|[\sigma]_k = \epsilon(d) f$ for all $\sigma \in \Gamma_0(N)$,

(iii) is meromorphic at all cusps.

If the form is holomorphic on $\mathcal{H}$ and at the cusps, then it is called a modular form. If it vanishes at the cusps, then it is called a cusp form. We now explain (iii).

Note that by (ii) $f(z+1) = f(z)$. Thus, f has a Fourier expansion in terms of $q = e^{2\pi i z}$, say $f = \sum a_n q^n$.

We say that f is meromorphic (respectively, holomorphic, vanishes) at ∞ if $a_n = 0$ for all $n <$ some n_0 (respectively

$n < 0$, $n \leq 0$). We say that f is meromorphic (respectively, holomorphic, vanishes) at the cusps if $f|[\sigma]_k$ is meromorphic (respectively, holomorphic, vanishes) at ∞ for all $\sigma \in SL_2(\mathbf{Z})$.

For example, if $N = 1$, then $\Gamma_0(N) = \Gamma_1(N) = SL_2(\mathbf{Z})$. If f is a modular form of this level, then its Nebentypus must be trivial. Moreover, $f|[\sigma]_k = f$ for all $\sigma \in SL_2(\mathbf{Z})$ and so the cusp condition only need be checked at ∞. In general, one has finitely many conditions to check, taking σ running through the finitely many cosets of $\Gamma_0(N)$ in $SL_2(\mathbf{Z})$.

The set of modular forms (respectively cusp forms) of weight k, level N, and Nebentypus ϵ will be denoted $M_k(N, \epsilon)$ (respectively $S_k(N, \epsilon)$). As will be shown later, these are finite-dimensional $\mathbf{C}$-vector spaces.

Exercise: Show that if f, f' are modular functions of weights k, k' respectively and level 1, then $ff', f/f'$ are modular functions of weights $k+k', k-k'$ respectively and level 1. Show that for $\lambda \in \mathbf{C}$, λf and if $k = k'$, then $f + f'$ are modular functions of weight k, level 1.

Example: Let Λ be the lattice in $\mathbf{C}$ generated by fundamental periods $1, \tau$, where $\tau \in \mathcal{H}$. $G_{2k}(\tau) = G_{2k}(\Lambda) = \sum_{(m,n)\neq(0,0)} \frac{1}{(m\tau+n)^{2k}}$ (the Eisenstein series) is a modular form of weight $2k$ and level 1. (ii) follows since $G_{2k}(\lambda\Lambda) = \lambda^{-2k} G_{2k}(\Lambda)$ and (iii) follows since uniform convergence allows passage to limit term by term, the $m \neq 0$ terms giving 0, the $m = 0$ terms giving $\sum_{n\neq 0} n^{-2k} = 2\zeta(2k)$.

For this same Λ, $\Delta = g_2^3 - 27g_3^2$ is therefore a modular form of weight 12 and level 1. Using the known values of $\zeta(4), \zeta(6)$, we get its constant coefficient 0, and so it is a cusp form. In fact, looking deeper shows that $\Delta = (2\pi)^{12} q \prod_{n=1}^{\infty}(1-q^n)^{24}$ [28].

Henceforth, we shall normalize Δ so that its first coefficient is 1.

Another important example is $j(\tau) := \frac{1728g_2^3}{\Delta}$, which is a modular function of weight 0 and level 1, and thus defines a map $j : SL_2(\mathbf{Z})\backslash\mathcal{H} \to \mathbf{C}$. Its Fourier expansion $\frac{1}{q} + 744 + 196884q + \ldots$ has fascinating connections with the Monster finite simple group.

The best way to think of modular forms is in terms of associated Riemann surfaces, called modular curves.

4.4.1 Riemann surfaces

A *surface* is a topological space S which is Hausdorff and connected such that there is an open cover $\{U_\alpha | \alpha \in A\}$ and homeomorphisms ϕ_α of U_α to open sets $V_\alpha \subseteq \mathbf{C}$. Then (U_α, ϕ_α) is called a *chart* and the set of charts an *atlas.* If $U_\alpha \cap U_\beta \neq \emptyset$, then transition function $t_{\alpha\beta} = \phi_\beta\phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$. Call S a *Riemann surface* if all the $t_{\alpha\beta}$, where defined, are analytic.

Example: (i) Let $\mathbf{C}_\infty = \mathbf{C} \cup \{\infty\}$. Take the topology with open sets those in $\mathbf{C}$ together with $\{\infty\} \cup (\mathbf{C} - K)$ (K compact in $\mathbf{C}$). Let $U_\alpha = \mathbf{C}, \phi_\alpha(z) = z$ and $U_\beta = \mathbf{C}_\infty - \{0\}, \phi_\beta(z) = 1/z(z \in \mathbf{C}), \phi_\beta(\infty) = 0$. These two charts make $\mathbf{C}_\infty$ a compact Riemann surface, identifiable with the sphere.

(ii) If Λ is a lattice in $\mathbf{C}$, then $\mathbf{C}/\Lambda$ is a compact Riemann surface.

We now introduce another useful class of compact Riemann surfaces:

Definition 4.18 *Let $\mathcal{H}^* = \mathcal{H} \cup \mathbf{Q} \cup \{\infty\}$, and put a topology on $\mathcal{H}^*$ by taking the following as basic open sets:*

(i) about a point in $\mathcal{H}$ any open disk entirely inside $\mathcal{H}$;

(ii) about ∞, $\{\mathrm{Im}\tau > r\}$ for any $r > 0$;

(iii) about $x \in \mathbf{Q}$, $D \cup \{x\}$, where D is any open disk in $\mathcal{H}$ of radius $y > 0$ and center $x + iy$.

Extend the action of $\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ on $\mathcal{H}$ to act on $x = [x, 1] \in \mathbf{Q}$ and $\infty = [1, 0]$ by $\sigma[x, y] = [ax + by, cx + dy] \in P^1(\mathbf{Q})$ (i.e. homogenized). Let $Y_i(N) = \Gamma_i(N)\backslash\mathcal{H} \subset \Gamma_i(N)\backslash\mathcal{H}^* = X_i(N)$. These are compact Riemann surfaces and called modular curves. See [19] p.311 or [11] p.76 for more details.

Riemann surfaces form a category where the morphisms are analytic maps defined thus.

Definition 4.19 *Call a continuous map $f : R \to S$ of Riemann surfaces analytic if the maps $\psi_\beta f \phi_\alpha^{-1}$ are analytic maps on domains in $\mathbf{C}$, wherever defined. An analytic map $f : R \to \mathbf{C}_\infty$ is a meromorphic function on R (matches the usual definition, setting $f(p) = \infty$, if f has a pole at p). The collection of all such functions is a field, the function field $K(R)$.*

Example: (i) The meromorphic functions on torus $\mathbf{C}/\Lambda$ are just the elliptic functions with respect to Λ (in fact $= \mathbf{C}(\wp, \wp')$).

(ii) The meromorphic functions on modular curve $X_0(N)$ are just the modular functions of weight 0, level N, and trivial Nebentypus.

The most important result here is the Open Mapping Theorem, stating that any analytic nonconstant map of Riemann surfaces maps open sets to open sets.

Exercise: Using this, show that if $f : R \to S$ is one such and R compact, then $f(R) = S$, and so S is compact. Deduce Liouville's theorem.

Moreover, show that f is a k-to-1 map for some k (hint: let $S_m \subseteq S$ be those points with precisely m preimages, counting multiplicity. Show that S_m is open and then use compactness and connectedness of S).

Lemma 4.20 *The function j is surjective.*

Proof: Consider j as a map $X_0(1) \to \mathbf{C}_\infty$. Since Δ has a simple zero at ∞ and no others (as we showed in establishing E_Λ is an elliptic curve), j has a simple pole at ∞ and no others. Thus, invoking the exercise above, j is 1-to-1. QED

Corollary 4.21 *Every elliptic curve over $\mathbf{C}$ is of the form E_Λ, for some lattice Λ.*

Proof: Given $y^2 = f(x)$ with $f \in \mathbf{C}[x]$ a cubic with distinct roots, we can, by change of variables, get $y^2 = 4x^3 - Ax - B$ for some $A, B \in \mathbf{C}$. The claim is that there exists Λ such that $g_2(\Lambda) = A, g_3(\Lambda) = B$. From the definition of j above, we can find τ such that $\frac{g_3^2}{g_2^3}$ takes any value other than $\frac{1}{27}$. Pick τ such that this equals $\frac{B^2}{A^3}$ ($\neq \frac{1}{27}$ since f has distinct roots). Choose λ such that $g_2(\tau) = \lambda^4 A$. Then $g_3(\tau)^2 = \lambda^{12}B^2$, so $g_3(\tau) = \pm\lambda^6 B$. If we have the negative sign, then replace λ by $i\lambda$. Noting that Eisenstein series satisfy by definition $g_2(\lambda L) = \lambda^{-4}g_2(L), g_3(\lambda L) = \lambda^{-6}g_3(L)$, we are done if we take $\Lambda = \lambda L$ where L has basis $1, \tau$. QED

Note that by the last theorem, j identifies $X_0(1)$ with $\mathbf{C} \cup \{\infty\}$. In fact:

Theorem 4.22 *The modular functions of weight 0 and level 1 are precisely the rational functions of j, i.e. the function field*

$K(X_0(1)) = \mathbf{C}(j)$.

Proof: An earlier exercise showed that rational functions in j are modular functions of that weight and level. Conversely, if f is such a function, say with poles τ_i, counted with multiplicity, then $g = f\prod_i(j(\tau) - j(\tau_i))$ is a modular function of weight 0 and level 1 with no poles in $\mathcal{H}$. If g has a pole of order n at ∞, then there exists c such that $g - cj^n$ has a pole of order $n-1$ at ∞ (and no others). By induction, g minus some polynomial in j has no pole in $\mathcal{H}^*$, and so is constant. Thus, g, and so $f \in \mathbf{C}(j)$. QED

Letting α_i run over all integer matrices

$$\Delta_N := \{\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \text{ with } ad = N, d > 0, 0 \le b < d, \gcd(a,b,d) = 1\}$$

(of which there are $\mu(N) := N\prod_{p|N}(1+\frac{1}{p})$ such matrices), the *modular polynomial* of order N is $\Phi_N(x) = \prod_{i=1}^{\mu(N)}(x - j\circ\alpha_i)$. One root is $j_N := j\circ\alpha$, where $\alpha = \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}$ (i.e. $j_N(z) = j(Nz)$).

In fact, $\mu(N) = [SL_2(\mathbf{Z}) : \Gamma_0(N)]$ and so is the degree of the cover $X_0(N) \to X_0(1)$. The corresponding extension of function fields $K(X_0(1)) \subseteq K(X_0(N))$ is then also of degree $\mu(N)$. $K(X_0(N))$ turns out to be $K(X_0(1))(j_N)$:

Theorem 4.23 *See [19], p. 336 on. $\Phi_N(x)$ has coefficients in $\mathbf{Z}[j]$ and is irreducible over $\mathbf{C}(j)$ (and so is the minimal polynomial of j_N over $\mathbf{C}(j)$). The function field $K(X_0(N)) = \mathbf{C}(j, j_N)$.*

This enables us to define $X_0(N)$, a priori a curve over $\mathbf{C}$, over $\mathbf{Q}$. This means that it can be given by equations over $\mathbf{Q}$. If $N > 3$, then one can further define a scheme over $\mathbf{Z}[1/N]$, so that base change via $\mathbf{Z}[1/N] \to \mathbf{Q}$ yields this curve (in other words, we have a model for $X_0(N)$ over $\mathbf{Q}$ with good reduction at primes not dividing N).

Proof: First, we note that j, j_N are indeed in $K(X_0(N))$, i.e. satisfy $f(z) = f(\sigma z)$ for all $\sigma \in \Gamma_0(N)$. This clearly holds for j since j is a modular function of weight 0 on all of $SL_2(\mathbf{Z})$. Since $j_N(\sigma z) = j \circ \alpha \circ \sigma(z) = j \circ (\alpha\sigma\alpha^{-1})\alpha(z)$ and $\alpha\sigma\alpha^{-1} = \begin{pmatrix} a & Nb \\ N^{-1}c & d \end{pmatrix} \in SL_2(\mathbf{Z})$, $j_N(\sigma z) = j \circ \alpha(z) = j_N(z)$. The condition at the cusps is easily checked.

Next, we note that $j \circ \alpha_i (1 \leq i \leq \mu(N))$ are distinct functions on $\mathcal{H}$. If $\alpha_i = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$, then $j \circ \alpha_i(z) = j(\frac{az+b}{d}) = \frac{1}{q_d^a \zeta_d^b} + \ldots$, where $q_d = e^{2\pi i z/d}, \zeta_d = e^{2\pi i/d}$. If $j \circ \alpha_i = j \circ \alpha_{i'}$, take the quotient and let $\mathrm{Im} z \to 0$ to get $q_d^a \zeta_d^b = q_{d'}^{a'} \zeta_{d'}^{b'}$. So $a/d = a'/d'$, but since $ad = N = a'd'$ and all are positive, $a = a', d = d'$, whence $b = b'$ too.

Next, we show the properties of Φ_N. If $\gamma \in SL_2(\mathbf{Z})$, then we check that $\alpha_i \gamma = \beta \alpha_k$ for some k and $\beta \in SL_2(\mathbf{Z})$. Thus $j \circ \alpha_i \circ \gamma = j \circ \alpha_k$, whence γ permutes the roots of Φ_N, and so its coefficients are invariant under $SL_2(\mathbf{Z})$, hence $\in \mathbf{C}(j)$ (meromorphic since polynomials in the $j \circ \alpha_i$). In fact, one easily computes that $SL_2(\mathbf{Z})$ acts transitively on the roots of Φ_N, whence Φ_N is irreducible over $\mathbf{C}(j)$. To show its coefficients lie in $\mathbf{Z}[j]$, we see that they lie in $\mathbf{Z}[\zeta_N]$ since $d \mid N$. Automorphism $\zeta_N \mapsto \zeta_N^r$ (any r coprime to N) permutes the $j \circ \alpha_i$, and so the coefficients lie in $\mathbf{Q} \cap \mathbf{Z}[\zeta_N] = \mathbf{Z}$.

QED

These polynomials Φ_N tend to have huge coefficients, but at least they define $X_0(N)$ as a curve over $\mathbf{Q}$.

If X is a compact Riemann surface of genus g and $W =$

$\Omega_{hol}(X)$ its holomorphic differentials, then W and so $V = \hom(W, \mathbf{C})$ are g-dimensional $\mathbf{C}$-vector spaces.

Exercise: Show that there is an isomorphism of $\mathbf{C}$-vector spaces given by

$$\Omega_{hol}(X_0(N)) \to S_2(N), \quad f(z)dz \mapsto f(z).$$

(Hint: if $f(z)$ is such a cusp form and $\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$, then $f(\sigma z)d(\sigma z) = (cz+d)^2 f(z)(cz+d)^{-2}dz = f(z)dz$. The holomorphicity of $f(z)dz$ corresponds to being a cusp form since $2\pi i dz = dq/q$.)

This shows that the dimension of $S_2(N)$ is the genus of $X_0(N)$ (in particular finite), computable by Riemann-Roch (explicitly given in [19]). Likewise, the finite dimensionality of $S_2(N, \epsilon)$ is bounded by the genus of $X_1(N)$. There are similar interpretations of $S_k(N)$ for higher k.

Let $C_1, ..., C_{2g}$ denote the usual $2g$ cycles on the g-handled X, generating free abelian group $H_1(X, \mathbf{Z})$. Let $\Lambda = \mathrm{Im}(H_1(X, \mathbf{Z}) \to \hom(W, \mathbf{C}))$ where the map is $C \mapsto (\omega \mapsto \int_C \omega)$, a discrete subgroup of rank $2g$ called the *period lattice* of X. The *Jacobian* of X, $\mathrm{Jac}(X)$, is the g-dimensional complex torus $\mathbf{C}^g/\Lambda$, isomorphic as a group to $(\mathbf{R}/\mathbf{Z})^{2g}$. This is an example of an *abelian variety* over $\mathbf{C}$.

The *Abel map* is $X \to Jac(X), x \mapsto \{\int_{x_0}^x \omega_j\}$ (for some fixed x_0 whose choice doesn't matter since the image is defined up to a period in Λ). In the case $g = 1$, then this is bijective. Let $Div(X)$ be the free abelian group on the points of X and $Div^0(X)$ its subgroup consisting of the elements whose coefficients sum to 0. The Abel map extends to a group homomorphism $Div(X) \to Jac(X)$, whose restriction to $Div^0(X)$ is surjective with kernel the so-called *principal divisors* $\mathcal{P}(\mathcal{X})$.

Since $Div^0(X_0(N))/\mathcal{P}(X_0(N))$ makes sense over $\mathbf{Q}$, this defines $J_0(N) := \mathrm{Jac}(X_0(N))$ over $\mathbf{Q}$ (in fact it is a coarse moduli scheme over $\mathbf{Z}[1/N]$ if $N > 3$). Galois action on its division points then produces Galois representations $G_{\mathbf{Q}} \to GL_{2g}(\mathbf{Z}/n)$. From this we obtain 2-dimensional Galois representations associated to certain modular forms, defined below.

Definition 4.24 *If f is a cusp form of weight k, level N, and Nebentypus ϵ, define the mth Hecke operator by*

$$T_m f = m^{(k/2)-1} \sum_j f|[\alpha_j]_k,$$

where if $N = 1$, the α_j run through Δ_n, and for general N through the matrices in Δ_n with $gcd(a, N) = 1$.

Exercise: If f has q-expansion at ∞, $\sum_{n=0}^{\infty} a_n q^n$, then $T_m f$ has q-expansion

$$\sum_{n=0}^{\infty} b_n q^n, \text{ where } b_n = \sum_{d|(m,n)} \epsilon(d) d^{k-1} a_{mn/d^2}$$

(taking $\epsilon(d) = 0$ if $(d, N) \neq 1$).

Show that $T_m f$ is again a cusp form of weight k, level N, and Nebentypus ϵ such that T_m is a linear operator on $S_{k,\epsilon}(N)$. Furthermore, check the Hecke operators commute.

Definition 4.25 *If $T_m f = \lambda_m f$ (some $\lambda_m \in \mathbf{C}$) for all m, then f is called a cuspidal eigenform. Actually, $a_m = \lambda_m a_1$ and we shall normalize eigenforms so that $a_1 = 1$ and so $\lambda_m = a_m$. The commutative ring the Hecke operators generate, $\mathcal{T}$, is called the Hecke algebra.*

For example, since $S_{12,1}(1)$ has dimension 1, basis Δ, Δ is a cuspidal eigenform. A very useful corollary to the definition follows:

Corollary 4.26 *If $f = \sum a_n q^n$ is a cuspidal eigenform, the map $T_m \mapsto a_m$ extends to a ring homomorphism (eigencharacter) $\theta : \mathcal{T} \to \mathbf{C}$.*

Consider the (injective) map $S_k(N) \to \mathbf{C}[[q]]$ that sends a cusp form to its Fourier expansion at ∞. Then the inverse image of $\mathbf{Z}[[q]]$ is $S_k(N;\mathbf{Z})$ and $S_k(N;A) = S_k(N;\mathbf{Z}) \otimes_{\mathbf{Z}} A$. Note that using the explicit coefficients of T_m, $\mathcal{T}$ acts on $S_k(N;\mathbf{Z})$. The q-expansion principle says that $S_k(N;\mathbf{C}) = S_k(N)$, i.e. $S_k(N)$ has a basis in $S_k(N;\mathbf{Z})$, and so $\mathcal{T}$ embeds in $\mathrm{End} S_k(N;\mathbf{Z})$. This has various consequences:

Theorem 4.27 *$\mathcal{T}$ is a finite free $\mathbf{Z}$-algebra. If f is a cuspidal eigenform, then there exists an algebraic number ring O_f containing all coefficients of f and the image of ϵ. (Note that the image of θ above lies in O_f.)*

If A is a ring, we let $S_k(N;A)$ denote the cusp forms of level N and weight k defined over A. Note that

$$S_k(N;A) = \mathrm{hom}_{A-alg}(\mathcal{T}, A). \quad (\dagger)$$

This follows for $A = \mathbf{Z}$ by mapping $\phi : S_k(N;\mathbf{Z}) \to \mathrm{hom}(\mathcal{T}, \mathbf{Z})$ by $f \mapsto (t \mapsto a_1(tf))$ and then noting that ϕ is injective and that $\mathcal{T}$ is free of rank $\leq$ that of $S_k(N;\mathbf{Z})$. The general case follows by tensoring with A.

Note that the Hecke operators T_m also act on $Div^0(X_0(N))$ (and so $J_0(N)$) by extending linearly $T_m[z] = \sum[\alpha_i z]$, where $[z]$ is the orbit of $z \in \mathcal{H}$ and α_i runs (again) through all matrices $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ with $ad = n, d > 0, gcd(a,N) = 1, 0 \leq b < d$. Let I_p

denote the image of inertia under $G_{\mathbf{Q}_p} \to G_{\mathbf{Q}}$, D_p the image of $G_{\mathbf{Q}_p}$, and Fr_p the element of D_p/I_p mapping to the Frobenius map in $G_{\mathbf{F}_p}$. For λ a prime ideal of O_f, denote the fraction field of $(O_f)_\lambda$ by K_λ.

Theorem 4.28 *Let $f = \sum a_n q^n$ be a cuspidal eigenform of weight k, level N, and Nebentypus ϵ. For each prime λ, there exists a unique semisimple continuous homomorphism $\rho : G_{\mathbf{Q}} \to GL_2(K_\lambda)$ such that if prime $p \not| \ell N$, then $\rho(I_p) = \{1\}, \mathrm{tr}\rho(Fr_p) = a_p, \det\rho(Fr_p) = \epsilon(p)p^{k-1}$.*

Proof: We explain the case $k = 2$ and trivial Nebentypus, since the elliptic curve representations we care about will correspond to this case.

Let ℓ be the rational prime below λ. Let $J_0(N)[\ell^n] \cong (\mathbf{Z}/\ell^n)^{2g}$ denote the kernel of multiplication by ℓ^n on $J_0(N)$ and $T_\ell(J_0(N)) = \varprojlim J_0(N)[\ell^n] \cong \mathbf{Z}_\ell^{2g}$ the corresponding Tate module, a free $\mathbf{Z}_\ell$-module of rank $2g$ on which $G_{\mathbf{Q}}$ acts (the argument for elliptic curves carries over to g-dimensional tori), where g is the genus of $X_0(N)$. The action of T_m on $J_0(N)$ given above carries over to $T_\ell(J_0(N))$. Then $W := T_\ell(J_0(N)) \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$ is a $\mathcal{T} \otimes_{\mathbf{Z}} \mathbf{Q}_\ell$-module, in fact free of rank 2. (The proof of this comes from the Hodge decomposition $S_2(N) \oplus \bar{S}_2(N) = H^1(X, \mathbf{C})$, where $\bar{S}_2(N)$ gives the anti-holomorphic differentials, and the fact that if A is a field of characteristic 0, then $S_k(N; A)$ is free of rank 1 over $\mathcal{T} \otimes_{\mathbf{Z}} A$ by (†) above.) This yields a representation $G_{\mathbf{Q}} \to GL_2(\mathcal{T} \otimes_{\mathbf{Z}} \mathbf{Q}_\ell)$. Note that the particular level N eigenform has not been used yet.

The eigencharacter $\mathcal{T} \to O_f$ now induces a homomorphism $\mathcal{T} \otimes_{\mathbf{Z}} \mathbf{Q}_\ell \to O_f \otimes_{\mathbf{Z}} \mathbf{Q}_\ell$. Since O_f is an order in a number field, $O_f \otimes_{\mathbf{Z}} \mathbf{Q}_\ell \cong \prod_\lambda K_\lambda$. Mapping onto the λth component yields $\rho :$

$G_{\mathbf{Q}} \to GL_2(K_\lambda)$. We shall see that ρ has the desired properties in the next chapter. QED

Exercise: Show that if O is the ring of integers in a number field, then the splitting of ℓ in O determines the form of $O \otimes_{\mathbf{Z}} \mathbf{Q}_\ell$. (Hint: $\mathbf{Z}[x]/(f(x)) \otimes_{\mathbf{Z}} \mathbf{Q}_\ell \cong \mathbf{Q}_\ell[x]/(f(x))$.)

For cases with $k > 2$, Deligne [7] used higher symmetric powers of the Tate module. For $k = 1$, Deligne and Serre [8] built the representation from congruent eigenforms of higher weight.

Exercise: Let $\rho : G_{\mathbf{Q}} \to GL_n(K)$ be a continuous homomorphism where K is a local field with valuation ring O. Considering K^n thus as a $G_{\mathbf{Q}}$-module, show that there is a stable lattice O^n under this action.

Letting O be the valuation ring of K_λ and F its residue field, the *residual representation* associated to a cuspidal eigenform f will denote the composition $G_{\mathbf{Q}} \xrightarrow{\rho} GL_2(O) \to GL_2(F)$ produced above.

The Big Picture. We have three sources of naturally occurring Galois representations, namely elliptic curves, modular forms, and étale group schemes. The most general is the last, and these will be used next to define the general class of *semistable* representations we shall focus on. We shall see that this class contains the representations coming from an elliptic curve associated by Frey to a putative counterexample to Fermat's Last Theorem. The first thing is to create links between the various kinds of representations defined.

5
Invariants of Galois representations, semistable representations

Let F be a finite field of characteristic $\ell > 2$ and V a 2-dimensional F-vector space on which $G_{\mathbf{Q}}$ acts continuously. Thus we have a representation $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(F)$. By composing ρ with the homomorphism $G_{\mathbf{Q}_\ell} \to G_{\mathbf{Q}}$, V has a $G_{\mathbf{Q}_\ell}$-action and so defines an étale group scheme over $\mathbf{Q}_\ell$.

Definition 5.1 *Call $\bar{\rho}$ good at ℓ if this group scheme comes via base change from a finite flat group scheme over $\mathbf{Z}_\ell$ and also* $\det\bar{\rho}|_{I_\ell} = \chi|_{I_\ell}$*, the cyclotomic character restricted to the inertia subgroup I_ℓ. Call $\bar{\rho}$ ordinary at ℓ if $\bar{\rho}|_{I_\ell}$ can be written $\begin{pmatrix} \chi_\ell|_{I_\ell} & * \\ 0 & 1 \end{pmatrix}$. Call $\bar{\rho}$ semistable at ℓ if it is good or ordinary at ℓ.*

Let prime $p \neq \ell$. Call $\bar{\rho}$ good at p if $\bar{\rho}$ is unramified at p, i.e. $\bar{\rho}(I_p) = \{1\}$. Call $\bar{\rho}$ semistable at p if $\bar{\rho}(I_p)$ is unipotent (i.e. all its eigenvalues are 1*). Call $\bar{\rho}$ semistable if it is semistable at all primes.*

5.1 Serre Invariants

Given $\bar{\rho}$ as above, we want to define positive integers $N(\bar{\rho}), k(\bar{\rho})$, and a group homomorphism $\epsilon(\bar{\rho}) : (\mathbf{Z}/N(\bar{\rho}))^{\times} \to \mathbf{C}^{\times}$. Suppose $\bar{\rho}$ is *odd*, i.e. $\det\bar{\rho}(c) = -1$, where c is complex conjugation, and that $\bar{\rho}$ is absolutely irreducible, i.e. $V \otimes_F \bar{F}$ has no $G_{\mathbf{Q}}$-invariant proper subspace. We shall consider the following two conjectures of Serre [30]:

Conjecture 5.2 *(Strong conjecture) There exists a cuspidal eigenform f of weight $k(\bar{\rho})$, level $N(\bar{\rho})$, and Nebentypus $\epsilon(\bar{\rho})$ whose associated residual representation is $\bar{\rho}$.*

Conjecture 5.3 *(Weak conjecture) There exists a cuspidal eigenform whose associated residual representation is $\bar{\rho}$.*

Work of Kenneth Ribet and others [24],[10] established that the weak conjecture implies the strong conjecture. $N(\bar{\rho})$ comes from the local representations $\bar{\rho}_p$ for $p \neq \ell$ and $k(\bar{\rho})$ from $\bar{\rho}_\ell$. We begin with the *Artin conductor* $N(\bar{\rho})$.

Consider $\bar{\rho}_p : G_{\mathbf{Q}_p} \to GL_2(F)$ $(p \neq \ell)$. Let $\mathcal{G}_i$ denote the (finite) image of the ith ramification subgroup (using upper numbering, in fact, [29]; note that $\mathcal{G}_0 = G_0$ and $\mathcal{G}_1 = G_1$). Let V_i denote the subspace of V fixed by $\mathcal{G}_i$, and set

$$n(p,\bar{\rho}) = \sum_{i=0}^{\infty} \frac{\dim(V/V_i)}{|\mathcal{G}_0/\mathcal{G}_i|}.$$

It is a deep theorem [29] that $n(p,\bar{\rho})$ is a non-negative integer. More easily, note that

$(i) n(p,\bar{\rho}) = 0 \iff V = V_0 \iff \bar{\rho}_p$ unramified, $(\bar{\rho}(\mathcal{G}_0) = \{1\})$

$(ii) n(p,\bar{\rho}) = \dim(V/V_0) \iff V = V_1 \iff \bar{\rho}_p$ tamely ramified, $(\bar{\rho}(\mathcal{G}_1) = \{1\})$

Definition 5.4 *The Artin conductor of $\bar{\rho}$ is defined by*

$$N(\bar{\rho}) = \prod_{p \neq \ell} p^{n(p,\bar{\rho})},$$

well-defined by (i) above, since only finitely many primes ramify under $\bar{\rho}$.

Since det$\bar{\rho} : G_{\mathbf{Q}} \rightarrow F^{\times}$ has abelian image, it factors through $Gal(L/\mathbf{Q})$ for some abelian extension $L/\mathbf{Q}$. By Kronecker-Weber, there is a smallest m such that $L \subseteq \mathbf{Q}(\zeta_m)$, and one can show that $m = \ell N(\bar{\rho})$. Thus, det gives a character $(\mathbf{Z}/m)^{\times} \cong (\mathbf{Z}/\ell)^{\times} \times (\mathbf{Z}/N(\bar{\rho}))^{\times} \rightarrow \mathbf{F}^{\times}$. Restricting this to the second factor yields homomorphism $\bar{\epsilon} : (\mathbf{Z}/N(\bar{\rho}))^{\times} \rightarrow F^{\times} \hookrightarrow \bar{F}^{\times}$. Lifting this to $\bar{\mathbf{Z}} \subset \mathbf{C}$ gives $\epsilon : (\mathbf{Z}/N(\bar{\rho}))^{\times} \rightarrow \mathbf{C}^{\times}$. This defines $\epsilon(\bar{\rho})$.

We would like to have det$\bar{\rho} = \epsilon(\bar{\rho})\chi^{k(\bar{\rho})-1}$, and this determines $k(\bar{\rho}) \pmod{\ell - 1}$. By the result at the end of chapter 3, $\bar{\rho}(\mathcal{G}_1) = \{1\}$. Thus $\bar{\rho}|_{I_\ell} : I_\ell \rightarrow GL_2(\bar{F})$ has cyclic image of order prime to ℓ, and so is diagonalizable. Let $\phi, \phi' : I_\ell \rightarrow \bar{F}^{\times}$ be the two characters this produces. The exact description of $k(\bar{\rho})$ is quite complicated but just depends on what ϕ, ϕ' are in terms of fundamental characters (see chapter 3) [30].

Theorem 5.5 *$\bar{\rho}$ is semistable if and only if $N(\bar{\rho})$ is squarefree, $\epsilon(\bar{\rho})$ is trivial, and $k(\bar{\rho}) = 2$ or $\ell + 1$.*

Proof: The form of $\epsilon(\bar{\rho})$ and of $k(\bar{\rho}) \pmod{\ell - 1}$ corresponds to det $\bar{\rho}$ being exactly the cyclotomic character. The exact form of $k(\bar{\rho})$ requires Serre's prescription above. As for $N(\bar{\rho})$ being squarefree, $n(p, \bar{\rho}) = 1$ if and only if V_0 is of dimension 1 and $V_1 = V$, so if and only if the representation is ordinary at p. QED

5.2 Fontaine-Mazur Conjecture

The notions of “good” and “semistable” extend to representations to rings other than fields. Suppose R is a complete, Noetherian local ring with residue field finite of characteristic ℓ. Then R is a $\mathbf{Z}_\ell$-algebra and we can view the cyclotomic character as mapping to R via $\chi : G_{\mathbf{Q}} \to \mathbf{Z}_\ell^\times \to R^\times$. Let $\rho : G_{\mathbf{Q}} \to GL_2(R)$ be a continuous homomorphism, giving a $G_{\mathbf{Q}}$-action on $V := R^2$.

Definition 5.6 *If $|R| < \infty$, then call ρ good at ℓ if the corresponding étale group scheme over $\mathbf{Q}_\ell$ comes from a finite flat group scheme over $\mathbf{Z}_\ell$ and det $\rho|_{I_\ell} = \chi|_{I_\ell}$. For general R, call ρ good at ℓ if for every (closed) ideal I of finite index in R, the representation $\rho : G_{\mathbf{Q}} \to GL_2(R/I)$ is good at ℓ. Call ρ ordinary at ℓ if there is a short exact sequence $0 \to V^{-1} \to V \to V^0 \to 0$ of free R-modules stable under $G_{\mathbf{Q}_\ell}$, such that I_ℓ acts trivially on V^0 and as χ on V^{-1} so that*

$$\rho|_{I_\ell} \cong \begin{pmatrix} \chi|_{I_\ell} & * \\ 0 & 1 \end{pmatrix}.$$

Call ρ semistable at ℓ if ρ is good or ordinary at ℓ and $\mathrm{det}\rho|_{I_\ell} = \chi|_{I_\ell}$.

The following is a special case of the major conjecture of Fontaine and Mazur [12] that says that every potentially semistable ℓ-adic Galois representation arises from the Galois action on some subquotient of an ℓ-adic cohomology group of a variety over $\mathbf{Q}$.

Conjecture 5.7 *(Fontaine-Mazur) If O is the valuation ring of a finite extension K of $\mathbf{Q}_\ell$, and $\rho : G_{\mathbf{Q}} \to GL_2(O)$ a continuous, absolutely irreducible homomorphism, unramified at all* *but finitely many primes, and semistable at ℓ, then there exists a cuspidal eigenform f whose associated Galois representation is ρ.*

Wiles' work amounts to establishing some special, nontrivial cases of this conjecture.

5.3 Reduction of elliptic curves

Our goals now are to introduce Frey's elliptic curves associated to a putative counterexample to Fermat's Last Theorem, to define semistable elliptic curves and show that Frey's curves are semistable, and to show that the Galois representations attached to semistable elliptic curves are semistable. First, we need a little on reduction of elliptic curves.

A *Weierstrass model* for an elliptic curve E over $\mathbf{Q}$ is an equation of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

By completing square and cube it can be written $y^2 = x^3 - 27c_4x - 54c_6$, where $c_4 = b_2^2 - 24b_4, c_6 = -b_2^3 + 36b_2b_4 - 216b_6$ and $b_2 = a_1^2 + 4a_2, b_4 = 2a_4 + a_1a_3, b_6 = a_3^2 + 4a_6$. This curve has discriminant $\Delta = (c_4^3 - c_6^2)/1728$.

Definition 5.8 *We say that E has good reduction over $\mathbf{Q}_p$ (or at p) if there is a Weierstrass equation for E with $a_i \in \mathbf{Z}_p$ that (when the coefficients are reduced modulo p) defines an elliptic curve over $\mathbf{F}_p$ (which happens if and only if the corresponding $\Delta \notin p\mathbf{Z}_p$).*

We have to be careful - for example, $y^2 = x^3 - 1$ and $y^2 = x^3 - 64$ define the same elliptic curve over $\mathbf{Q}$ but with the reductions mod 2 being nonsingular and having a cusp. Another example is the *Frey elliptic curve* with equation $y^2 = x(x-A)(x+B)$, where $A, B, C := -A-B$ are nonzero relatively prime integers. We shall focus particularly on the case $A = a^\ell, B = b^\ell, C = c^\ell$, with $\ell \geq 5$ a prime and $gcd(a,b,c) = 1$ (i.e. a counterexample to $(FLT)_\ell$). For this equation, $\Delta = 16(ABC)^2$ and so if $p \not| 2ABC$, then E has good reduction at p. For the $(FLT)_\ell$ situation, however, we may assume, by relabeling, that $A \equiv -1 \pmod 4$ and $B \equiv 0 \pmod{32}$ and then changing variables by $x \mapsto 4x, y \mapsto 8y + 4x$ yields equation

$$y^2 + xy = x^3 + \frac{B-A-1}{4}x^2 - \frac{AB}{16}x.$$

For this equation, $\Delta = 2^{-8}(ABC)^2$. Moreover, $c_4 = A^2 + AB + B^2$, which is checked to be relatively prime to abc, i.e. $v_p(c_4) = 0$ for all primes p of bad reduction. This is what is called a *minimal Weierstrass model* for the curve, in that $|\Delta|$ is now as small as possible, say Δ_{min}.

Exercise: Considering how changes of variable of the form $x \mapsto u^2x, y \mapsto u^3y$ affect Δ, c_4, c_6, show that a Weierstrass model is minimal if $v_p(\Delta) < 12$ or $v_p(c_4) < 4$ or $v_p(c_6) < 6$ for every prime p. (Here v_p denotes p-adic valuation.) Show that changes of variable do not change the j-invariant of the equation. Deduce that two elliptic curves are isomorphic over an algebraically closed field if and only if they have the same j-invariant.

So, for our example, $v_p(c_4) = 0$ gives that the model is minimal. Thus, if $p|abc$, the Frey curve necessarily has bad reduction at p. In fact, its reduction mod p has a *node*.

Definition 5.9 *Suppose p is a prime of bad reduction for E. E*

mod p has a unique singular point. If that point is a node (respectively cusp), then we say E has multiplicative (respectively additive) reduction at p. A node (respectively cusp) means that two (respectively three) of the roots of the cubic are equal.

Say E has semistable reduction at p if it has good or multiplicative reduction at p. Call E semistable if its reduction at all primes is semistable.

Multiplicative reduction occurs at p if and only if $v_p(c_4) = 0, v_p(\Delta) > 0$. (We see this directly for the Frey curves for odd primes p since if p divides a, it does not divide b and so only two roots of $x(x - a^\ell)(x + b^\ell)$ coalesce modulo p.) Thus, Frey elliptic curves are semistable.

What is particularly nice about semistable elliptic curves is that their associated Galois representations are semistable. This is established via the theory of Tate curves:

5.3.1 Tate elliptic curves

Consider first the complex case, where Λ is the lattice in $\mathbf{C}$ with basis $1, \tau (\in \mathcal{H})$. There is a group isomorphism $\mathbf{C}/\mathbf{Z} \to \mathbf{C}^\times$ given by $x \mapsto e^{2\pi i x}$. Setting $q = e^{2\pi i \tau}$, this yields a group isomorphism $\mathbf{C}/\Lambda \to \mathbf{C}^\times/q^{\mathbf{Z}}$, where $q^{\mathbf{Z}}$ is the infinite cyclic group with generator q. We thus have a multiplicative representation of $E_\Lambda(\mathbf{C})$.

Explicitly, in terms of q, setting $\frac{1}{(2\pi i)^2} x = X + \frac{1}{12}$ and $\frac{1}{(2\pi i)^3} y = 2Y + X$, the equation for E_Λ becomes:

$$y^2 + xy = x^3 + a_4(q) + a_6(q), a_4(q) = -5s_3(q), a_6(q) = -(5s_3(q) + 7s_5(q))/12,$$

and $s_k(q) = \sum_{n=1}^{\infty} \frac{n^k q^n}{1-q^n}$. Note that $a_4(q), a_6(q) \in \mathbf{Z}[[q]]$, so this defines an elliptic curve over $\mathbf{Z}[[q]]$ since the discriminant of this curve is $q \prod_{n=1}^{\infty} (1 - q^n)^{24}$ and we get its j-invariant as $\frac{1}{q} +$

$744 + 196884q + \ldots (*)$, as usual. Tate observed that much of this carries over from $\mathbf{C}$ to p-adic local fields:

Theorem 5.10 *(Tate) Let $K/\mathbf{Q}_p$ be a finite extension with valuation v and absolute value $|u| = c^{v(u)}$ for some $0 < c < 1$. Suppose $q \in K^\times$ satisfies $|q| < 1$. Then $s_k(q), a_4(q), a_6(q)$ are defined as above and converge in K. Then $y^2 + xy = x^3 + a_4(q) + a_6(q)$ defines an elliptic curve E_q over K and for any algebraic extension L/K, $L^\times/q^{\mathbf{Z}} \cong E_q(L)$.*

Since $|q| < 1$, the reduction of the curve is $y^2 + xy = x^3$, which has bad multiplicative reduction (in fact split - meaning that the tangent lines at the node are defined over the residue field).

Conversely, if E has multiplicative reduction at p, then its j-invariant has $v_p(j) < 0$ and the series $q = j^{-1} + 744j^{-2} + 750420j^{-3} + ...$, inverting $(*)$, converges to give q such that $v_p(q) = -v_p(j) > 0$, so $|q| < 1$. The corresponding Tate elliptic curve E_q is isomorphic over $\bar{\mathbf{Q}}_p$ (in fact over a quadratic extension of $\mathbf{Q}_p$) to E since they have the same j-invariant. The division points of E_q are particularly simple to study, whence the action of $G_{\mathbf{Q}_p}$ on them is easily given.

Theorem 5.11 *Let E have multiplicative reduction at p. Then there exists $q \in \mathbf{Q}_p^\times$ such that $E(\bar{\mathbf{Q}}_p) \cong (\bar{\mathbf{Q}}_p^\times/q^{\mathbf{Z}})(\delta)$ as $G_{\mathbf{Q}_p}$-modules, where δ is the unique unramified quadratic character, i.e. $\sigma(\phi(x)) = \phi(\sigma(x^{\delta(\sigma)}))$ for all $\sigma \in G_{\mathbf{Q}_p}$, where $\delta : G_{\mathbf{Q}_p} \to \{\pm 1\}$ is trivial if the reduction is split, and is the unique unramified quadratic character otherwise. Thus,*

$$\rho_{E,\ell^\infty}|_{G_{\mathbf{Q}_p}} \cong \begin{pmatrix} \chi_\ell & * \\ 0 & 1 \end{pmatrix} \otimes \delta.$$

Proof: $E_q[\ell^n] \subseteq E_q(\bar{\mathbf{Q}}_p)$ and if $x \in \bar{\mathbf{Q}}_p^\times/q^{\mathbf{Z}}$, x^{ℓ^n} is trivial if and only if $x^{\ell^n} = q^m$ for some $m \in \mathbf{Z}$, so if and only if $x = \zeta_{\ell^n}^a (q^{1/\ell^n})^b$ for some $a, b \in \mathbf{Z}/\ell^n$. Consider the map

$$E_q[\ell^n] \to < q > / < q^{\ell^n} > \cong \mathbf{Z}/\ell^n, x \mapsto x^{\ell^n},$$

a homomorphism with kernel μ_{ℓ^n}. $G_{\mathbf{Q}_p}$ acts via the cyclotomic character on the kernel and trivial on the image since $q \in \mathbf{Q}_p$. QED

Theorem 5.12 *If E is a semistable elliptic curve over $\mathbf{Q}$, then $\rho_{E,\ell^\infty} : G_{\mathbf{Q}} \to GL_2(\mathbf{Z}_\ell)$ is semistable.*

Proof: First, the Weil pairing gives that $\Lambda^2 T_\ell(E) \cong \mu_{\ell^\infty}$, as Galois modules, and so $\det\bar{\rho}$ is the cyclotomic character.

Second, if E has good reduction at ℓ, then for each n $E[\ell^n]$ comes from a finite flat group scheme over $\mathbf{Z}_\ell$, namely the minimal Weierstrass model. The representing $\mathbf{Z}_\ell$-algebra $\Re$ is free, since tensoring with $\mathbf{Q}_\ell$ and $\mathbf{F}_\ell$ gives the same rank and the only such finitely generated $\mathbf{Z}_\ell$-modules are free. Then ρ_{E,ℓ^∞} is good at ℓ.

If E has bad so multiplicative reduction at ℓ, then the previous theorem gives that the ℓ-adic representation is ordinary (note $\delta|_{I_\ell}$ is trivial).

Third, if E has good reduction at $p \neq \ell$, then by Néron-Ogg-Shafarevich below $E[\ell^n]$ is unramified at p, so ρ_{E,ℓ^∞} is good at p. If E has bad so multiplicative reduction at $p \neq \ell$, the previous theorem again applies. QED

Theorem 5.13 *(Néron-Ogg-Shafarevich) Suppose $p \not| m$. Then E has good reduction at p if and only if the $G_{\mathbf{Q}}$-action on $E[m]$ is unramified at p.*

Proof: We show the forwards direction, which is the one we need. Let K be a finite extension of $\mathbf{Q}_p$ such that $E[m] \subseteq E(K)$. Let F be the residue field of K. One checks that the kernel of reduction $\phi : E(K) \to E(F)$ has no points of order prime to p. Thus the restriction $E[m] \to E(F)$ is injective. Let $P \in E[m]$ with reduction $\tilde{P}, \sigma \in I_p$. Then $\phi(\sigma(P) - P) = \sigma(\tilde{P}) - \tilde{P} = 0$, and so by injectivity $\sigma(P) = P$. Thus, I_p acts trivially on $E[m]$. QED

This result is likewise true for abelian varieties, in particular $J_0(N)$, and thus shows that the representations at the end of chapter 4 are unramified at primes not dividing ℓN. Moreover, in both cases, the action on Fr_p on the m-division points is given by their action on their reductions.

Theorem 5.14 *Suppose E, an elliptic curve over $\mathbf{Q}$, has multiplicative reduction at $p > 2$, which may or may not be ℓ. Then $\rho_{E,\ell}$ is good at p if and only if $\ell | v_p(\Delta_{min})$.*

Proof: Consider the ℓ-division field of E_q, namely $K = \mathbf{Q}_p(\zeta_\ell, q^{1/\ell})$. If $p \neq \ell$, note that $\ell | v_p(\Delta_{min}) \iff \ell | v_p(q) \iff K/\mathbf{Q}_p$unramified $\iff$ $\rho_{E,\ell}$unramified at p. For $p = \ell$, need the argument in Edixhoven's article in [ ]. QED

For the Frey curves related to $(FLT)_\ell$, $\Delta_{min} = 2^{-8}(abc)^{2\ell}$ and so ℓ divides $v_p(\Delta_{min})$ for all odd primes p. By the last criterion $\rho_{E,\ell}$ is good at every prime except possibly 2. Being good at ℓ gives $k(\rho_{E,\ell}) = 2$, whereas being good at the other odd primes and semistable at 2 gives $N(\rho_{E,\ell}) = 2$. Since its determinant is just the cyclotomic with no twist, $\epsilon(\rho_{E,\ell})$ is trivial. Thus, if $\rho_{E,\ell}$ were associated to a cuspidal eigenform, then it would come from $S_2(2) = \{0\}$, a contradiction.

Moreover, Mazur's work [ ] shows that $\rho_{E,\ell}$ is absolutely irreducible and so a counterexample to $(FLT)_\ell$ would violate Serre's strong conjecture. The idea for this is that if say $\rho_{E,\ell}$ were reducible, then $E[\ell]$ would have a subgroup or quotient of order ℓ on which $G_{\mathbf{Q}}$ acts trivially. Such a thing corresponds to a rational point on $X_0(\ell)$, but Mazur found all of them and there are no noncuspidal ones for large ℓ.

Ribet [24], which inspired Wiles to begin his long journey to the proof, took the approach of trying to prove that Serre's weak conjecture implies his strong conjecture. This then shows that a counterexample to $(FLT)_\ell$ would violate $\rho_{E,\ell}$ being associated to *any* kind of cuspidal eigenform, i.e. the Shimura-Taniyama conjecture. Ribet, Diamond, and others established the full "weak implies strong" conclusion later [ ], but Ribet at first did enough for our purposes.

Theorem 5.15 *(Ribet) Suppose $\rho : G_{\mathbf{Q}} \to GL_2(F)$ is an odd, continuous, absolutely irreducible representation over a finite field F of characteristic $\ell > 2$. Suppose ρ is associated to some cuspidal eigenform of level N, weight 2, and trivial Nebentypus. If ρ is good at $p||N$, then ρ is associated to some cuspidal eigenform of level N/p whenever $p \not\equiv 1 \pmod{\ell}$ or $gcd(\ell, N) = 1$.*

A description of Ribet's approach will be given in an appendix.

Corollary 5.16 *The Shimura-Taniyama conjecture, that every Galois representation attached to an elliptic curve is attached to a modular form, implies Fermat's Last Theorem.*

Suppose we have a counterexample to $(FLT)_\ell, \ell \geq 5$. If the Shimura-Taniyama conjecture holds, then the associated Frey curve is associated to some cuspidal eigenform of squarefree level N. It is a theorem of Mazur that says that the corresponding ℓ-division representation is absolutely irreducible. By repeated use of Ribet's theorem, this representation is associated to some cuspidal eigenform of level 2 (weight 2 and trivial Nebentypus), but there are no such.

The Big Picture. The proof of Fermat's Last Theorem is hereby reduced to proving that all semistable elliptic curves are modular. This will be achieved by showing that certain families of semistable Galois representations are all associated to modular forms. We therefore need some way of parametrizing such families and this will be provided by universal (semistable/modular) Galois representations.

6
Deformation theory of Galois representations

A good description of this material can be found in [14]. The original source is [20].

Fix a profinite group G, a finite field F, and a continuous representation $\bar{\rho} : G \to GL_n(F)$. Let $\mathbf{C}_F$ denote the category whose objects are complete, Noetherian local (commutative) rings R with residue field $R/\mathbf{m}_R \cong F$ and whose morphisms are ring homomorphisms $\phi : R \to S$ such that $\phi(\mathbf{m}_R) \subseteq \mathbf{m}_S$ and such that the induced map $R/\mathbf{m}_R \to S/\mathbf{m}_S$ is the identity map on F. Cohen's theorem [2] says that every ring in $\mathbf{C}_F$ is a quotient ring of $W(F)[[T_1, ..., T_m]]$ for some m and so is a $W(F)$-algebra.

Definition 6.1 *A lift of $\bar{\rho}$ to a ring R in $\mathbf{C}_F$ is a continuous homomorphism $\rho : G \to GL_n(R)$ which modulo $\mathbf{m}_R$ produces $\bar{\rho}$. Two such lifts ρ_1 and ρ_2 are called strictly equivalent if there exists $M \in \Gamma_n(R) := ker(GL_n(R) \to GL_n(F))$ such that $\rho_2 = M^{-1}\rho_1 M$. A deformation of $\bar{\rho}$ is a strict equivalence class of* *lifts.*

Let $E(R) = \{\text{deformations of } \bar{\rho} \text{ to } R\}$. A morphism $R \to S$ induces a map of sets $E(R) \to E(S)$, making E a functor $\mathbf{C}_F - - \to \mathit{Sets}$. We establish conditions under which E is representable, so that there is one representation parametrizing all lifts of $\bar{\rho}$.

Say that G satisfies (†) if the maximal elementary ℓ-abelian quotient of every open subgroup of G is finite. Examples include:

(1) $G_{\mathbf{Q}_p}$ for any p;

(2) let S be a finite set of primes of $\mathbf{Q}$ and $G_{\mathbf{Q},S}$ denote the Galois group of the maximal extension of $\mathbf{Q}$ unramified outside S, i.e. the quotient of $G_{\mathbf{Q}}$ by the normal subgroup generated by the inertia subgroups I_p for $p \notin S$.

These satisfy (†) by class field theory. Note that by Néron-Ogg-Shafarevich a Galois representations associated to an elliptic curves or a modular form factors through $G_{\mathbf{Q},S}$ for some S.

Theorem 6.2 *(Mazur) Suppose that $\bar{\rho}$ is absolutely irreducible and that G satisfies* (†). *Then the associated functor E is representable.*

That means that there is a ring $\Re(\bar{\rho})$ in $\mathbf{C}_F$ and a continuous homomorphism $\xi : G \to GL_n(\Re(\bar{\rho}))$ such that all lifts of $\bar{\rho}$ to R in $\mathbf{C}_F$ arise, up to strict equivalence, via a unique morphism $\Re(\bar{\rho}) \to R$. The representing ring $R(\bar{\rho})$ is called the *universal deformation ring* of $\bar{\rho}$ and (the strict equivalence class of) ξ the *universal deformation* of $\bar{\rho}$. For example, if $n = 2$, $\bar{\rho}$ is odd, $F = \mathbf{Z}/\ell$, and $G = G_{\mathbf{Q},S}$ with $\ell \in S$, then $\Re(\bar{\rho})$ is typically $\mathbf{Z}_\ell[[T_1, T_2, T_3]]$.

One proof of Mazur's theorem follows Schlessinger's criteria [27], which were developed to study local singularities. We let $\mathbf{C}_F^0$ denote the subcategory of $\mathbf{C}_F$ consisting of Artinian rings. Note that if R is in $\mathbf{C}_F$, then every $R/\mathbf{m}_R^i$ is in $\mathbf{C}_F^0$. The *dual numbers* $F[\epsilon] = F[T]/(T^2)$ with ϵ the image of T. A morphism $R \rightarrow S$ is called *small* if it is surjective with kernel a principal ideal whose product with $\mathbf{m}_R$ is 0, for example the map $\pi : F[\epsilon] \rightarrow F$ given by $a + b\epsilon \mapsto a$.

Definition 6.3 *Suppose $E : \mathbf{C}_F \dashrightarrow$ Sets is a functor satisfying $|E(F)| = 1$. Let $R_1 \rightarrow R_0$ and $R_2 \rightarrow R_0$ be morphisms in $\mathbf{C}_F^0$. Consider the natural map*

$$(*)\ E(R_1 \times_{R_0} R_2) \rightarrow E(R_1) \times_{E(R_0)} E(R_2),$$

which exists since $E(R_1) \times_{E(R_0)} E(R_2)$ is the terminal object in the category of sets

$$\begin{tikzcd}
& S \arrow[dl] \arrow[dr] & \\
E(R_1) \arrow[dr] & & E(R_2) \arrow[dl] \\
& E(R_0) &
\end{tikzcd}$$

Schlessinger's criteria are as follows:

H1. *$R_2 \rightarrow R_0$ small implies $(*)$ surjective.*

H2. *If $R_0 = F, R_2 = F[\epsilon]$, and $R_2 \rightarrow R_0$ is the map π above, then $(*)$ is bijective.*

Note: If **H2** *holds, then $t_E := E(F[\epsilon])$ has an F-vector space structure (the tangent space of E).*

H3. *t_E is a finite-dimensional F-vector space.*

H4. *If $R_1 = R_2$ and $R_i \rightarrow R_0 (i = 1, 2)$ is the same small map, then $(*)$ is bijective.*

Theorem 6.4 *(Schlessinger)* **H1, H2, H3, H4** *hold if and only if E is representable.*

Proof: So to prove Mazur's result, we need to show that his functor E satisfies **H1, H2, H3, H4**. Let E_i denote the set of continuous homomorphisms $G \to GL_n(R_i)$ lifting $\bar{\rho}$. Let $K_i = \Gamma_n(R_i)$. Then $E(R_i) = E_i/K_i$. Let $R_3 = R_1 \times_{R_0} R_2$. We are interested in the map

$$(*)\ E_3/K_3 \to E_1/K_1 \times_{E_0/K_0} E_2/K_2,$$

when $R_2 \to R_0$ is small.

To show $(*)$ is surjective then, we take $\rho_1 \in E_1, \rho_2 \in E_2$, yielding the same element of E_0/K_0, i.e. $\bar{\rho}_1 = M^{-1}\bar{\rho}_2 M$ for some $M \in K_0$. Since $R_2 \to R_0$ is surjective, so is $K_2 \to K_0$. If $N \in K_2$ maps to M, then $(\rho_1, N^{-1}\rho_2 N)$ gives the desired element of E_3.

Exercise: $(*)$ is injective if $C_{K_2}(\rho_2(G)) \to C_{K_0}(\rho_0(G))$ is surjective.

If $\bar{\rho}$ is absolutely irreducible, then both these groups consist of scalar matrices and so this holds, ensuring **H4** holds. This actually follows from Nakayama's Lemma, since e.g. $R_2[G] \to End_{R_2}(R_2^n)$ is surjective after tensoring with F thanks to the absolute irreducibility of $\bar{\rho}$. If $R_0 = F$, then $K_0 = \{1\}$ and so the condition holds, whence **H2** follows.

As for **H3**, note that $\Gamma_n(F[\epsilon])$ is isomorphic to the direct product of n^2 copies of F^+, so is an elementary abelian ℓ-group. Thus any lift $\rho : G \to GL_n(F[\epsilon])$ of $\bar{\rho}$ factors through G/H, where $ker(\bar{\rho})/H$ is the maximal elementary abelian quotient of $ker(\bar{\rho})$. By (†), G/H is finite and so there are finitely many lifts to $F[\epsilon]$, proving **H3**. QED

We shall be interested in families of representations satisfying some further condition, such as semistability. For this we need Ramakrishna's refinement of Mazur's result.

Let X be a property of $W(F)[G]$-modules of finite cardinality which is closed under isomorphism, direct sums, taking submodules, and quotienting. Fix $\bar{\rho} : G \to GL_n(F)$ such that F^n considered as a $W(F)[G]$-module via $\bar{\rho}$ satisfies X. For $R \in \mathbf{C}_F^0$ (which must then have finite cardinality), let $E_X(R)$ denote the set of deformations in $E(R)$ satisfying X.

Theorem 6.5 *(Ramakrishna) E_X is a functor on $\mathbf{C}_F^0$. Moreover, if E satisfies* **H1, H2, H3, H4**, *then so does E_X (in which case both functors are representable, where E_X is extended to the category $\mathbf{C}_F$ by $E_X(R) = \varprojlim E_X(R/\mathbf{m}_R^i)$).*

Proof: Let R, S be objects in $\mathbf{C}_F^0$ and $\phi : R \to S$ a morphism. To show E_X a functor, we need to show that if $\rho : G \to GL_n(R)$ has X, then the composition map $G \to GL_n(S)$ also has X. This follows since if $B = R^n$ and $D = S^n$ both with the given G-action, then ϕ induces $B \to D$ making D a finitely generated (it's finite!) B-module, say a quotient of B^m - since having X is closed under direct product and quotient, B and so B^m and so D all have X.

The next thing to note is that **H1** for E_X implies **H2,H3,H4** too. This follows since restrictions of injective maps to subsets are still injective. This gives injectivity in **H2** and **H4** with **H1** giving surjectivity. As for **H3**, since $E_X(F[\epsilon]) \subseteq E(F[\epsilon])$, the tangent space of E_X is also finite-dimensional.

To prove **H1** for E_X, set $R_3 = R_1 \times_{R_0} R_2$. Let $\rho_1 \times_{\rho_0} \rho_2 \in E_X(R_1) \times_{E_X(R_0)} E_X(R_2)$. By **H1** for E, we get $\rho \in E(R_3)$ mapping to this element. We just need to show that ρ has X. Well,

$R_3 \hookrightarrow R_1 \times R_2$ induces $R_3^n \hookrightarrow R_1^n \times R_2^n$, making R_3^n a submodule of a direct product of $W(F)[G]$-modules with X, whence R_3^n with this G-action has X. QED

Theorem 6.6 *Suppose that E and E_X satisfy the hypotheses of the previous theorem. Let $\Re$ and $\Re_X$ be the respective deformation rings. Then there is a natural surjection $\Re \to \Re_X$.*

Proof: Let $\xi : G \to GL_n(\Re)$ and $\xi_X : G \to GL_n(\Re_X)$ denote the universal deformations. Since ξ_X is a lift of $\bar{\rho}$ and ξ parametrizes all such, there is a (unique) morphism $\phi : R \to R_X$ which after composition with ξ yields ξ_X. Let the image of ϕ be S. We thereby get a representation $\rho : G \to GL_n(S)$, which is of type X. By universality of $\Re_X$ we get a unique morphism $\Re_X \to S$ producing ρ. The composition $\Re_X \to S \hookrightarrow \Re_X$, by universality again, has to be the identity map, so $S = \Re_X$. QED

We shall show that semistability defines such a property X. Namely, fix a finite field F of characteristic $\ell > 2$ and a continuous homomorphism $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(F)$ which is absolutely irreducible and semistable. (Note that semistability includes that det $\bar{\rho}$ be the cyclotomic character and this makes $\bar{\rho}$ odd, so we need not make this an extra condition.) Fix a finite set Σ of rational primes. Let R be in $\mathbf{C}_F^0$. If $\rho : G_{\mathbf{Q}} \to GL_2(R)$ is a lift of $\bar{\rho}$, we say that ρ is of *type* Σ if

(1) det ρ is the cyclotomic character;

(2) ρ is semistable at ℓ;

(3) if $\ell \notin \Sigma$ and $\bar{\rho}$ is good at ℓ, then ρ is good at ℓ;

if $p \notin \Sigma \cup \{\ell\}$ and $\bar{\rho}$ is unramified at p, then ρ is unramified at p;

if $p \notin \Sigma \cup \{\ell\}$ and $\bar{\rho}$ is ramified (so ordinary) at p, then ρ is ordinary at p.

Note: (1) If E is a semistable elliptic curve over $\mathbf{Q}$ and $\rho_{E,\ell}$ is absolutely irreducible, then ρ_{E,ℓ^∞} is a lift of type Σ if Σ contains all the primes of bad reduction for E.

(2) If ρ is of type $\Sigma \subseteq \Sigma'$, then ρ is of type Σ'.

(3) If ρ is of type Σ, then ρ is unramified outside $\{p : p|\ell N(\bar{\rho})\} \cup \Sigma$.

(4) A lift ρ of $\bar{\rho}$, unramified outside Σ, with $det(\rho)$ the cyclotomic character, semistable at ℓ, is of type $\Sigma \cup \{\ell\}$.

Theorem 6.7 *Given a continuous absolutely irreducible homomorphism $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(F)$ and given Σ, there exists an object $\Re_\Sigma$ in $\mathbf{C}_F$ and a continuous homomorphism $\rho_\Sigma : G_{\mathbf{Q}} \to GL_2(\Re_\Sigma)$ such that every lift ρ of $\bar{\rho}$ to R in $\mathbf{C}_F$, of type Σ, is strictly equivalent to the composition of ρ_Σ with some ϕ, a unique morphism from $\Re_\Sigma$ to R.*

The proof proceeds by dealing with each of the conditions required for type Σ in turn. First, we deal with the group scheme condition. To apply Ramakrishna's criterion we consider the étale group scheme over $\mathbf{Q}_\ell$ corresponding to $G_{\mathbf{Q}_\ell} \to GL_n(R)$ with $Rin\mathbf{C}_F^0$ and say that ρ has property X if this group scheme extends to a finite flat group scheme over $\mathbf{Z}_\ell$. We just need to show that X is preserved under direct sums, sub, and quotient.

Definition 6.8 *If G is an affine group scheme over $\mathbf{Z}_\ell$, say represented by A, let G_{gen} be its fibre (or base change) over $\mathbf{Q}_\ell$, so represented by $A \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$. Let H_{gen} be a closed subgroup scheme of G_{gen}, say represented by $(A \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell)/J$ where J is a Hopf ideal (ideals that ensure the quotient is still a Hopf algebra). Let $I = \phi^{-1}(J)$ under $\phi : A \to A \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$. Then the group scheme H represented by A/I, a subgroup scheme of G, is called the schematic closure of H_{gen}.*

Lemma 6.9 *A/I is torsion-free.*

Proof: ϕ induces an injection $A/I \hookrightarrow (A \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell)/J$, which is torsion-free since it is a $\mathbf{Q}_\ell$-algebra. QED

It follows that H is a finite flat group scheme over $\mathbf{Z}_\ell$. This is why goodness is preserved under sub. As for quotient group schemes, if H and G are group schemes over R with $H(A)$ a normal subgroup of $G(A)$ for all R-algebras A, then typically $F(A) = G(A)/H(A)$ does not give a representable functor F on R-algebras. Raynaud showed how to fix this. Let $A' = \{a \in A | \mu(a) \equiv 1 \otimes a \pmod{I \otimes A})\}$. Then A' is considered as representing G/H. Finally, as regards direct sums, if g, H are finite flat group schemes over $\mathbf{Z}_\ell$ represented by R, S respectively, let $F(A) = G(A) \times H(A)$ for every $\mathbf{Z}_\ell$-algebra A. Then check that F is an affine group scheme represented by $R \otimes_{\mathbf{Z}_\ell} S$, which is a finite flat $\mathbf{Z}_\ell$-algebra.

The second kind of condition is of the following nature. Let G be a profinite group and I a closed subgroup. Call $\rho : G \to GL_2(R)$ I-ordinary if the fixed points of R^2 under I form a free direct summand of rank 1.

Theorem 6.10 *If $\bar{\rho}$ is I-ordinary and absolutely irreducible and G satisfies* (†)*, then it has a universal I-ordinary lift.*

Proof: Again, we need only prove **H1**. Let $E_I(R)$ denote the set of I-ordinary deformations of $\bar{\rho}$ to R. Let $R_3 = R_1 \times_{R_0} R_2$ and consider $\rho_1 \times_{\rho_0} \rho_2 \in E_I(R_1) \times_{E_I(R_0)} E_I(R_2)$. Then since Mazur's functor E is representable, we get a $\rho_3 \in E(R_3)$ mapping to $\rho_1 \times_{\rho_0} \rho_2$. It remains to show that ρ_3 is I-ordinary. QED

The third kind of condition is just that ρ have determinant the cyclotomic character. Without imposing this condition, we so far have a universal deformation $\xi : G_{\mathbf{Q}} \to GL_2(\Re)$ satisfying everything else contained in type Σ. In particular, for $p \notin S$, ξ is unramified at p. Thus, $\xi(Fr_p)$ (for $p \notin S$) are defined. Let $det\xi(Fr_p) = r_p \in \Re$ and let I be the ideal of $\Re$ generated by $r_p - p$ for $p \notin S$. The representation $\tilde{\xi} = \xi \pmod{I} : G_{\mathbf{Q}} \to GL_2(\Re/I)$ now has $det\tilde{\xi}(Fr_p) = r_p = p = \chi(Fr_p)$ for $p \notin S$. By Chebotarev's density theorem, the $Fr_p (p \notin S)$ are dense in $G_{\mathbf{Q},S}$ and so $det\tilde{\xi} = \chi$.

Incidentally, imposing all these conditions has actually reduced us to a ring $\Re/I$ which will turn out to be finite over $\mathbf{Z}_\ell$.

Finally, we need another description of the tangent space of our functors. Suppose E is a representable functor on $\mathbf{C}_F$, represented by $\Re$. Set $t_E = E(F[\epsilon])$, the *tangent space* of E, a finite-dimensional F-vector space. This dimension will turn out to be a useful invariant of both E and $\Re$, calculated via Galois cohomology.

Note that $t_E = \hom(\Re, F[\epsilon])$. Under every such morphism, $\mathbf{m}_{\Re}$ maps to $\{b\epsilon | b \in F\}$ with kernel containing $\mathbf{m}_{\Re}^2 + \ell\Re$, such that t_E is the dual space of $\mathbf{m}_{\Re}/(\mathbf{m}_{\Re}^2 + \ell\Re)$. For example, if $\Re = W(F)[[T_1, ..., T_r]]$, then $\mathbf{m}_{\Re} = (T_1, ..., T_r, \ell)$ and so $\dim_F t_E = \dim_F(\mathbf{m}_{\Re}/(\mathbf{m}_{\Re}^2 + \ell\Re)) = r$. Furthermore, if I is an ideal of $\Re$ such that $I \subseteq \mathbf{m}_{\Re}^2 + \ell\Re$, then the tangent space of $\Re/I$ is also r-dimensional.

The Big Picture. We have seen that given a semistable $\bar{\rho}$ and a finite set of primes Σ, there is a universal lift $\xi_\Sigma : G_{\mathbf{Q}} \to GL_2(\Re_\Sigma)$ which parametrizes all lifts of $\bar{\rho}$ that are semistable at ℓ and that have no worse ramification at $p \notin \Sigma$ than $\bar{\rho}$ has. A very rough idea of how big $\Re_\Sigma$ is, can be given by the dimension of its tangent space. We shall shortly describe a universal lift of $\bar{\rho}$ of type Σ parametrizing those lifts associated to modular forms and describe a useful invariant for measuring how large that representing ring. Some commutative algebra will then establish how an inequality between these two invariants suffices for showing that the two universal lifts (type Σ and modular of type Σ) coincide. First, we need to describe how to compute the tangential dimension of $\mathfrak{R}_\Sigma$ in terms of Galois cohomology.

7

Introduction to Galois cohomology

A good introductory text for this chapter is [35]. For more advanced material, see [21].

Fix $\bar{\rho} : G \to GL_n(F)$. Consider the possible lifts σ of $\bar{\rho}$ to $GL_n(F[\epsilon])$. In particular, there is the trivial lift, which we shall also denote $\bar{\rho}$, arising from the embedding $F \to F[\epsilon]$. Then $\sigma(g) = (1 + \epsilon a(g))\bar{\rho}(g)$ for some map $a : G \to M_n(F)$. The fact that σ is a group homomorphism, i.e. $\sigma(gh) = \sigma(g)\sigma(h)$, translates into $a(gh) = a(g) + \bar{\rho}(g)^{-1}a(h)\bar{\rho}(g)$.

Definition 7.1 *Let M be a G-module (where the actions are continuous if G is profinite). A 1-cocycle is a (continuous) map $f : G \to M$ such that $f(gh) = f(g) + gf(h)$ for all $g, h \in G$. A 1-coboundary is a (continuous) map $f : G \to M$ given by $f(g) = gx - x$ for some (fixed) $x \in M$. These are 1-cocycles and we shall set $H^1(G, M)$ to denote the quotient group.*

Exercise: Check the 1-cocycles and 1-coboundaries do indeed form abelian groups, the first containing the second. Show that if G acts trivially on M, then $H^1(G, M) = Hom(G, M)$.

Note that $GL_n(F)$ acts on $M_n(F)$ by conjugation (the *adjoint* action). Thus, the map $\bar{\rho}$ makes $M_n(F)$ into a G-module, to be denoted $\mathrm{Ad}(\bar{\rho})$. As noted in the first paragraph above, every lift of $\bar{\rho}$ to $F[\epsilon]$ produces a 1-cocycle. One checks that if two lifts are strictly equivalent, then their 1-cocycles differ by a 1-coboundary. This yields a map from $E(F[\epsilon])$ (deformations to the dual numbers) to $H^1(G, \mathrm{Ad}(\bar{\rho}))$. Conversely, given a 1-cocycle $a : G \to \mathrm{Ad}(\bar{\rho})$, defining $\sigma(g) = (1 + \epsilon a(g))\bar{\rho}(g)$ gives a lift of $\bar{\rho}$ to $F[\epsilon]$. In this way, we get a bijection

$$E(F[\epsilon]) \cong H^1(G, \mathrm{Ad}(\bar{\rho})).$$

Note that this gives an alternative way of seeing that $E(F[\epsilon])$ is an F-vector space. If we impose some property X and look at the deformations of $\bar{\rho}$ that have X, i.e. $E_X(F[\epsilon])$, this corresponds to some subgroup of $H^1(G, \mathrm{Ad}(\bar{\rho}))$ which we denote $H^1_X(G, \mathrm{Ad}(\bar{\rho}))$. In particular, if we start with a semistable, absolutely irreducible representation $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(F)$ (F of odd characteristic ℓ) and consider lifts of type Σ, then the subgroup will be denoted $H^1_\Sigma(G_{\mathbf{Q}}, \mathrm{Ad}(\bar{\rho}))$.

We want to identify this set by considering what sort of 1-cocycles correspond to the properties involved in being of type Σ.

First, we impose the condition that the lift σ should have determinant the cyclotomic character χ, which is the same as the determinant of $\bar{\rho}$. Then $\det(1 + \epsilon a(g)) = 1$ for all $g \in G_{\mathbf{Q}}$. Since

$$1 + \epsilon a(g) = \begin{pmatrix} 1 + \epsilon a_{11} & \epsilon a_{12} \\ \epsilon a_{21} & 1 + \epsilon a_{22} \end{pmatrix},$$

and $\epsilon^2 = 0$, this gives $1+\epsilon(a_{11}+a_{22}) = 1$, implying $\mathrm{trace}(a(g)) = 0$. The $G_{\mathbf{Q}}$-submodule of $\mathrm{Ad}(\bar{\rho})$ consisting of trace 0 matrices will be denoted $\mathrm{Ad}^0(\bar{\rho})$. We are therefore only interested in subgroups of $H^1(G_{\mathbf{Q}}, \mathrm{Ad}^0(\bar{\rho}))$.

Second, if H is a subgroup of G and M a G-module, then we can restrict 1-cocyles and 1-coboundaries from G to H, thus producing a restriction map

$$res : H^1(G, M) \to H^1(H, M).$$

More generally, any homomorphism $H \to G$ will produce such a restriction map. The local behavior corresponding to type Σ will be captured by the restrictions from $G_{\mathbf{Q}}$ to $G_{\mathbf{Q}_p}$ and/or I_p. For example, suppose a lift σ corresponds to an element of the kernel from $H^1(G_{\mathbf{Q}}, \mathrm{Ad}^0(\bar{\rho})) \to H^1(I_p, \mathrm{Ad}^0(\bar{\rho}))$. Then $\sigma(g) = \bar{\rho}(g)$ for all $g \in I_p$, so that σ is unramified at p if and only if $\bar{\rho}$ is unramified at p. Being in the kernel in fact ensures that the ramification at p of σ is no worse than that of $\bar{\rho}$.

Definition 7.2 *Let $M = \mathrm{Ad}^0(\bar{\rho})$. By local conditions we mean a collection $\mathcal{L} = \{L_p\}$, where for each prime p (including infinity) we are given a subgroup $L_p \leq H^1(G_{\mathbf{Q}_p}, M)$ such that for all but finitely many p $L_p = ker(H^1(G_{\mathbf{Q}_p}, M) \to H^1(I_p, M))$. (These are called the unramified classes and will be denoted $H^1_{ur}(G_{\mathbf{Q}_p}, M)$.) The corresponding Selmer group will be*

$$H^1_{\mathcal{L}}(G_{\mathbf{Q}}, M) = \{c \in H^1(G_{\mathbf{Q}}, M) : res_p(c) \in L_p \text{for all } p\},$$

where $res_p : H^1(G_{\mathbf{Q}}, M) \to H^1(G_{\mathbf{Q}_p}, M)$ is restriction.

Note that by $\mathbf{Q}_\infty$, we mean the real numbers, and so $G_{\mathbf{Q}_\infty} = Gal(\mathbf{C}/\mathbf{R})$ of order 2, identified as the subgroup of order 2 of $G_{\mathbf{Q}}$ generated by complex conjugation. Since F has odd characteristic, $H^1(G_{\mathbf{Q}_\infty}, M) = \{0\}$ by the following, and so L_∞ has to be $\{0\}$.

Example: Suppose G has order 2 and M odd order. Show that $H^1(G, M) = \{0\}$.

Proof: Let $G = \{1, c\}$. A 1-cocycle is a map $f : G \to M$ such that $f(gh) = f(g) + gf(h)$ for $g, h \in G$. Setting $g = 1$ gives $f(1) = 0$. Setting $g = h = c$ gives $0 = f(c) + cf(c)$. So $c \in M^- := \{m \in M | cm = -m\}$. Conversely, if $m \in M^-$, defining f by $f(1) = 0, f(c) = m$ gives a 1-cocycle.

If f is a 1-coboundary, then $f(g) = gx - x$ for some $x \in M$. In particular, $f(c) = cx - x$. Consider the map $\phi : M \to M^-$ given by $\phi(x) = cx - x$. Let $M^+ = \{m \in M | cm = m\}$. Since $|M|$ is odd, $M = M^+ \oplus M^-(m = \frac{m+cm}{2} + \frac{m-cm}{2})$. Then $|Im(\phi)| = |M|/|Ker(\phi)| = |M|/|M^+| = |M^-|$ and so ϕ is surjective. QED

Exercise: If G is infinite (pro)cyclic, generated say by g and M is finite, show that $H^1(G, M) = M/((g-1)M)$. [Hint: if $m \in M$, define $f : G \to M$ by $f(g^i) = m + gm + g^2m + ... + g^{i-1}m$. Show that f is continuous, i.e. there exists N such that if $i \equiv j \pmod N$, then $f(g^i) = f(g^j)$, and satisfies the 1-cocycle condition. Conversely, if f is a 1-cocycle, use the 1-cocycle condition to obtain the form of $f(g^i)$ by induction on i.]

Theorem 7.3 *$H^1_\Sigma(G_\mathbf{Q}, \mathrm{Ad}(\bar\rho)) = H^1_\mathcal{L}(G_\mathbf{Q}, M)$, where $M = \mathrm{Ad}^0(\bar\rho)$ and $\mathcal{L}$ is given by:*

$L_\infty = 0,$
$L_p = H^1_{ur}(G_{\mathbf{Q}_p}, M)$ *if* $p \notin \Sigma \cup \{\ell\}$,
$= H^1(G_{\mathbf{Q}_p}, M)$ *if* $p \in \Sigma, p \neq \ell$,
$= H^1_f(G_{\mathbf{Q}_p}, M)$ *if* $p = \ell \notin \Sigma$,
$= H^1_{ss}(G_{\mathbf{Q}_p}, M)$ *if* $p = \ell \in \Sigma$.

Here, H^1_f and H^1_{ss} are the flat and semistable cohomology groups respectively, to be defined below.

This theorem is simply a restatement of what it means for a lift to be of type Σ - all we do is translate the conditions across to cohomology. Our main aim will be to compute the size of this Selmer group, giving the tangential dimension of $\Re_\Sigma$. First, since the property of $\bar{\rho}$ being good or ordinary is defined in terms of the $G_{\mathbf{Q}}$-module F^2, we need another interpretation of H^1.

Definition 7.4 *Given $\bar{\rho} : G \to GL_n(F)$, consider $V = F^n$, a G-module thanks to $\bar{\rho}$, and consider the set of extensions E of V by V, consisting of short exact sequences*

$$0 \to V \xrightarrow{\alpha} E \xrightarrow{\beta} V \to 0$$

of $F[G]$-modules. Call two such extensions equivalent if there is an isomorphism $i : E_1 \to E_2$ making the following diagram commute

$$\begin{tikzcd}
0 \arrow[r] & V \arrow[r] \arrow[d, "1_V"] & E_1 \arrow[r] \arrow[d, "i"] & V \arrow[r] \arrow[d, "1_V"] & 0 \\
0 \arrow[r] & V \arrow[r] & E_2 \arrow[r] & V \arrow[r] & 0
\end{tikzcd}$$

Let $Ext^1_{F[G]}(V,V)$ denote the set of equivalence classes.

Theorem 7.5 *There is a bijection between $H^1(G, \mathrm{Ad}(\bar{\rho}))$ and $Ext^1_{F[G]}(V,V)$.*

Proof: Pick $\phi : V \to E$ such that $\beta(\phi(m)) = m$ for all $m \in V$. Given $g \in G$, define $T_g : V \to V$ by

$$m \mapsto \alpha^{-1}(g\phi(g^{-1}m) - \phi(m)).$$

T_g can be considered as an n by n matrix. One checks that $T_{gh} = T_g + gT_h$, where gT_h means the matrix T_h conjugated by $\bar{\rho}(g)$. One also checks that equivalent extensions correspond to 1-cocycles that differ by a 1-coboundary. QED

Recall that an $F[G_{\mathbf{Q}_\ell}]$-module V of finite cardinality is *good* if there is a finite, flat group scheme H over $\mathbf{Z}_\ell$ such that $V \cong H(\bar{\mathbf{Q}}_\ell)$ as $F[G_{\mathbf{Q}_\ell}]$-modules. Recall that V is called *ordinary* if there is an exact sequence

$$0 \to V^{-1} \to V \to V^0 \to 0$$

of $F[G_{\mathbf{Q}_\ell}]$-modules such that I_ℓ acts trivially on V^0 and via the cyclotomic character on V^{-1}. V is called *semistable* if it is good or ordinary (or both).

Definition 7.6 *Identify* $H^1(G_{\mathbf{Q}_\ell}, \mathrm{Ad}(\bar{\rho}))$ *with* $Ext^1_{F[G_{\mathbf{Q}_\ell}]}(V, V)$. *Let* $H^1_{ss}(G_{\mathbf{Q}_\ell}, \mathrm{Ad}(\bar{\rho}))$ *consist of those extensions of* V *by* V *that are semistable. If* $\bar{\rho}$ *is not good at* ℓ*, take* $H^1_f(G_{\mathbf{Q}_\ell}, \mathrm{Ad}(\bar{\rho}))$ *to be* H^1_{ss}*. If* $\bar{\rho}$ *is good at* ℓ*, take* $H^1_f(G_{\mathbf{Q}_\ell}, \mathrm{Ad}(\bar{\rho}))$ *to consist of those extensions of* V *by* V *that are good.* $H^1_{ss}(G_{\mathbf{Q}_\ell}, \mathrm{Ad}^0(\bar{\rho}))$ *and* $H^1_f(G_{\mathbf{Q}_\ell}, \mathrm{Ad}^0(\bar{\rho}))$ *are defined by intersecting the above subgroups with* $H^1(G_{\mathbf{Q}_\ell}, \mathrm{Ad}^0(\bar{\rho}))$.

Theorem 7.7 $H^1_{\mathcal{L}}(G_{\mathbf{Q}}, \mathrm{Ad}^0(\bar{\rho}))$ *is a finite group.*

Note: we actually already know this theorem since its dimension over F is the tangential dimension of $\mathfrak{R}_\Sigma$, but it is good preparation for finding the exact order of the Selmer group..

Proof: Let $M = \mathrm{Ad}^0(\bar{\rho})$. Let S be a finite set of primes containing all the "bad" ones, i.e. ∞, ℓ, the primes p such that I_p acts nontrivially on M, and such that $L_p \neq H^1_{ur}$. By definition of Selmer group, there is an exact sequence

$$0 \to H^1_{\mathcal{L}}(G_{\mathbf{Q}}, M) \to H^1(G_{\mathbf{Q},S}, M) \to \oplus_{p\in S} H^1(G_{\mathbf{Q}_p}, M)/L_p.$$

Let $H = ker(G_{\mathbf{Q},S} \to Aut(M))$, the kernel of the action of $G_{\mathbf{Q},S}$ on M. Since M is finite, H is an open, finite index, normal subgroup of $G_{\mathbf{Q},S}$. By an above exercise, since H acts trivially on M, $H^1(H, M) = Hom(H, M)$. This is a finite group since the Hermite-Minkowski theorem says that there are only finitely many extensions of degree ℓ of the fixed field of H (a finite extension of $\mathbf{Q}$) unramified outside the primes above S.

The inflation-restriction exact sequence (see below) says that

$$0 \to H^1(G_{\mathbf{Q},S}/H, M^H) \to H^1(G_{\mathbf{Q},S}, M) \to H^1(H, M)$$

is exact, where $M^H := \{m \in M | gm = m \text{for all} g \in H\}$. Since $G_{\mathbf{Q},S}/H$ and M^H are finite, so is $H^1(G_{\mathbf{Q},S}/H, M^H)$ and so is $H^1(H, M)$ as noted above. Thus, so is $H^1(G_{\mathbf{Q},S}, M)$ and so is its subgroup $H^1_{\mathcal{L}}(G_{\mathbf{Q}}, M)$. QED

The next theorem identifies the kernel of the restriction maps introduced above.

Theorem 7.8 *(Inflation-Restriction) If M is a G-module and H is a normal subgroup of G, then M^H is a G/H-module and there is an exact sequence:*

$$0 \to H^1(G/H, M^H) \xrightarrow{\text{inf}} H^1(G, M) \xrightarrow{\text{res}} H^1(H, M).$$

Corollary 7.9 $H^1_{ur}(G_{\mathbf{Q}_p}, M) = H^1(G_{\mathbf{Q}_p}/I_p, M^{I_p})$

The advantage of working with cohomology groups is that they fit into various exact sequences and perfect pairings, allowing them to be computed in terms of simpler objects. One such simpler object is $H^0(G,M)$, which is defined to be M^G. An example of this sort of simplification is:

Theorem 7.10 *For finite M, $|H^1_{ur}(G_{\mathbf{Q}_p},M)| = |H^0(G_{\mathbf{Q}_p},M)| < \infty$.*

Proof: Consider

$$0 \to M^{G_{\mathbf{Q}_p}} \to M^{I_p} \overset{Fr_p - 1}{\longrightarrow} M^{I_p} \to M^{I_p}/((Fr_p - 1)M^{I_p}) \to 0$$

and by the last exercise,

$$H^1(G_{\mathbf{Q}_p}/I_p, M^{I_p}) = H^1(< Fr_p >, M^{I_p}) = M^{I_p}/((Fr_p - 1)M^{I_p}).$$

QED

One can define 2-cocycles and 2-coboundaries to be certain maps $f : G \times G \to M$. Namely, a 2-cocycle f is a map that satisfies

$$gf(h,k) - f(gh,k) + f(g,hk) - f(g,h) = 0,$$

whereas a 2-coboundary is an f of the form

$$f(g,h) = gF(h) - F(gh) + F(g)$$

for some $F : G \to M$. Then take $H^2(G,M) = \{2-\text{cocycles}\}/\{2-\text{coboundaries}\}$, and in fact we can likewise define r-cocycles and r-coboundaries.

Theorem 7.11 *Suppose $0 \to A \to B \to C \to 0$ is a short exact sequence of G-modules. Then there is a long exact sequence*

$$\begin{gathered} 0 \to H^0(G,A) \to H^0(G,B) \to H^0(G,C) \to \\ H^1(G,A) \to H^1(G,B) \to H^1(G,C) \to \\ H^2(G,A) \to H^2(G,B) \to H^2(G,C) \dots \end{gathered}$$

This sequence continues with groups $H^i(G, A)$ defined for all integers $i \geq 0$.

Before giving Wiles' big result on sizes of Selmer groups, we need some preparatory results. Note that if A and B are G-modules, then $\hom(A, B)$ has a G-action given by $(g(f))(a) = g(f(g^{-1}a))$ where $g \in G, f \in \hom(A, B), a \in A$.

Theorem 7.12 *(Local Tate duality) Suppose M is a $G_{\mathbf{Q}_p}$-module of finite cardinality, n. Set $M^* = \hom(M, \mu_n(\bar{\mathbf{Q}}_p))$, where $\mu_n(\bar{\mathbf{Q}}_p)$ is the nth roots of 1 in $\bar{\mathbf{Q}}_p$ given its natural $G_{\mathbf{Q}_p}$-action.*

There is a nondegenerate pairing for $i = 0, 1, 2$

$$H^i(G_{\mathbf{Q}_p}, M) \times H^{2-i}(G_{\mathbf{Q}_p}, M^*) \to H^2(\mathcal{G}_{\mathbf{Q}_p}, \mu_n) \hookrightarrow \mathbf{Q}/\mathbf{Z}.$$

If p does not divide the order of M, then under the pairing $H^1_{ur}(G_{\mathbf{Q}_p}, M)$ and $H^1_{ur}(G_{\mathbf{Q}_p}, M^*)$ are the exact annihilators of each other.

Theorem 7.13 *(Wiles) Let $\mathcal{L} = \{L_p\}$ be local conditions and $\mathcal{L}^* = \{L_p^*\}$, where $L_p^* \subseteq H^1(G_{\mathbf{Q}_p}, M^*)$ is the annihilator of L_p under Tate's pairing. By the last remark, these are also local conditions.*

$$\frac{|H^1_{\mathcal{L}}(G_{\mathbf{Q}}, M)|}{|H^1_{\mathcal{L}^*}(G_{\mathbf{Q}}, M^*)|} = \frac{|H^0(G_{\mathbf{Q}}, M)|}{|H^0(G_{\mathbf{Q}}, M^*)|} \prod_{p \leq \infty} \frac{|L_p|}{|H^0(G_{\mathbf{Q}_p}, M)|}$$

Note that since for all but finitely many primes $L_p = H^1_{ur}(G_{\mathbf{Q}_p}, M)$, which has the same order as $H^0(G_{\mathbf{Q}_p}, M)$, the righthandside is a finite product.

Proof: Here's the idea. The Poitou-Tate 9-term sequence (see e.g. Milne's book) and the definition of the Selmer group yield exact sequence (S being the set of bad primes introduced earlier and c complex conjugation):

$$0 \to H^0(G_{\mathbf{Q},S}, M) \to (\oplus_{p\in S} H^0(G_{\mathbf{Q}_p}, M))/((1+c)M) \to H^2(G_{\mathbf{Q},S}, M^*)^\wedge \to$$

$$H^1_{\mathcal{L}}(G_{\mathbf{Q}}, M) \to \oplus_{p\in S} L_p \to H^1(G_{\mathbf{Q},S}, M^*)^\wedge \to H^1_{\mathcal{L}^*}(G_{\mathbf{Q}}, M^*)^\wedge \to 0$$

Here $A^\wedge$ denote the dual $\mathrm{hom}(A, \mathbf{Q}/\mathbf{Z})$.

The alternating product of orders of groups in an exact sequence is 1, and so

$$\frac{|H^1_{\mathcal{L}}(G_{\mathbf{Q}}, M)|}{|H^1_{\mathcal{L}^*}(G_{\mathbf{Q}}, M^*)|} = \frac{|H^0(G_{\mathbf{Q},S}, M)||H^2(G_{\mathbf{Q},S}, M^*)||(1+c)M|\prod_{p\in S}|L_p|}{|H^1(G_{\mathbf{Q},S}, M^*)|\prod_{p\in S}|H^0(G_{Q_p}, M)|}.$$

Since we are happiest computing orders of H^0's (or if necessary H^1's), we need some way of removing the H^2 term here, which is provided by the *global Euler characteristic formula*:

$$\frac{|H^1(G_{\mathbf{Q},S}, M^*)|}{|H^0(G_{\mathbf{Q},S}, M^*)||H^2(G_{\mathbf{Q},S}, M^*)|} = \frac{|M^*|}{|H^0(G_{\mathbf{Q}_\infty}, M^*)|} = |(1+c)M|.$$

QED

Note that this last module is what was called M^+ in a previous example. There is also a *local Euler characteristic formula*:

Theorem 7.14 *If M is a finite $G_{\mathbf{Q}_p}$-module, then*

$$\frac{|H^1(G_{\mathbf{Q}_p}, M)|}{|H^0(G_{\mathbf{Q}_p}, M)||H^2(G_{\mathbf{Q}_p}, M)|} = p^{v_p(|M|)}.$$

This will be needed in studying how H^1_Σ varies as we vary Σ. Suppose, for instance, we add a prime into Σ, yielding Σ'. Then $\Re'_\Sigma$ maps onto $\Re_\Sigma$. We can control the difference in their tangential dimensions as follows.

Theorem 7.15 *Suppose $|M|$ is a power of ℓ and $q \neq \ell$ a prime for which $L_q = H^1_{ur}(G_{\mathbf{Q}_q}, M)$. Define $\mathcal{L}'$ by $L'_p = L_p$ if $p \neq q$, and $L'_q = H^1(G_{\mathbf{Q}_q}, M)$. (This corresponds to $\Sigma' = \Sigma \cup \{q\}$.) Then*

$$\frac{|H^1_{\mathcal{L}'}(G_{\mathbf{Q}}, M)|}{|H^1_{\mathcal{L}}(G_{\mathbf{Q}}, M)|} \leq |H^0(G_{\mathbf{Q}_q}, M^*)|.$$

Proof: If $\mathcal{L}$ is replaced by $\mathcal{L}'$, consider what happens to the terms on the righthandside of Wiles' formula. They remain the same except for the term for $p = q$, which changes from 1 to $|H^1(G_{\mathbf{Q}_q}, M)|/|H^0(G_{\mathbf{Q}_q}, M)| = |H^2(G_{\mathbf{Q}_q}, M)|$, using the above Euler characteristic formula. By the local Tate duality pairing, $|H^2(G_{\mathbf{Q}_q}, M)| = |H^0(G_{\mathbf{Q}_q}, M^*)|$.

We must also consider the effect on $\mathcal{L}^*$. Let $\mathcal{L}'^*$ denote the new dual local conditions. Since $L'^*_q = 0$, the conditions defining $H^1_{\mathcal{L}'^*}$ are more restrictive than those defining $|H^1_{\mathcal{L}^*}$. Thus, $|H^1_{\mathcal{L}'^*}(G_{\mathbf{Q}}, M^*)| \leq |H^1_{\mathcal{L}^*}(G_{\mathbf{Q}}, M^*)|$.

Putting this together,

$$\frac{|H^1_{\mathcal{L}'}(G_{\mathbf{Q}}, M)|}{|H^1_{\mathcal{L}}(G_{\mathbf{Q}}, M)|} = \frac{|H^1_{\mathcal{L}'^*}(G_{\mathbf{Q}}, M^*)|}{|H^1_{\mathcal{L}^*}(G_{\mathbf{Q}}, M^*)|}|H^0(G_{\mathbf{Q}_q}, M^*)| \leq |H^0(G_{\mathbf{Q}_q}, M^*)|.$$

QED

The remaining matter is to compute the individual terms on the righthandside of Wiles' formula. Let us start with $p = \infty$. $H^0(G_{\mathbf{Q}_\infty}, M) = M^{G_{\mathbf{Q}_\infty}}$. We are assuming $\bar{\rho}$ is odd so can take the image of complex conjugation to be $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and so with $M = \mathrm{Ad}^0(\bar{\rho})$, we seek those trace 0 matrices fixed under conjugation by that matrix, easily computed to be $\{\begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix} |a \in F\}$. $|L_\infty| = 1$ and so the $p = \infty$ term contributes $1/|F|$.

Suppose $p \neq \ell$. If $p \notin \Sigma$, we already noted that $|L_p| = |H^0(G_{\mathbf{Q}_p}, M)|$ and so the ratio in the formula is 1 for such primes.

If $p \neq \ell$ and $p \in \Sigma$, then $|L_p|/|H^0(G_{\mathbf{Q}_p}, M)| = |H^1(G_{\mathbf{Q}_p}, M)|/|H^0(G_{\mathbf{Q}_p}, M)| = |H^2(G_{\mathbf{Q}_p}, M)|$ (by local Euler characteristic formula) $= |H^0(G_{\mathbf{Q}_p}, M^*)|$ (by local Tate duality).

Thus, the only hard part lies in computing the $p = \ell$ contribution to the formula. (Note that the "global" terms, e.g. $|H^0(G_{\mathbf{Q}}, M)|$ will typically be 1, since we assume that $\bar{\rho}$ is absolutely irreducible, whence the centralizer of M consists of scalar matrices, but the only such of trace 0 is the zero matrix.)

In the case of H^1_f, the theory of Fontaine and Lafaille is used, whereby they work with a category equivalent to that of finite flat $W(F)[G_{\mathbf{Q}_\ell}]$-modules, namely whose objects are $W(F)$-modules D of finite cardinality together with a distinguished submodule D^0 and $W(F)$-linear maps $\phi_{-1} : D \to D$ and $\phi_0 : D \to D$ satisfying $\phi_{-1}|_{D^0} = \ell\phi_0$ and $Im(\phi_{-1}) + Im(\phi_0) = D$. This reduces the computations to linear algebra, and it turns out that

$$\frac{|L_\ell|}{|H^0(G_{\mathbf{Q}_\ell}, \mathrm{Ad}^0(\bar{\rho}))|} = |F|.$$

As for H^1_{ss}, if we let $W \leq \mathrm{Ad}^0(\bar{\rho}) = M$ be $\{\begin{pmatrix} 0 & * \\ 0 & 0 \end{pmatrix}\}$, then this is a $G_{\mathbf{Q}_\ell}$-submodule since $\bar{\rho}$ restricted to $G_{\mathbf{Q}_\ell}$ has the form (taking a suitable basis) $\begin{pmatrix} \chi\psi_1 & * \\ 0 & \psi_2 \end{pmatrix}$. Here the ordinariness of $\bar{\rho}$ at ℓ translates into ψ_1, ψ_2 being unramified characters (i.e. trivial on I_ℓ). Since

$$\begin{pmatrix} r & s \\ 0 & t \end{pmatrix}\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}\begin{pmatrix} r & s \\ 0 & t \end{pmatrix}^{-1} = \begin{pmatrix} 0 & ra/t \\ 0 & 0 \end{pmatrix},$$

the $G_{\mathbf{Q}_\ell}$-action on W is via the character $\chi\psi_1/\psi_2$ (†).

If we want to consider lifts to $F[\epsilon]$ that are ordinary at ℓ too, then we need to set

$$L_\ell := ker(H^1(G_{\mathbf{Q}_\ell}, M) \to H^1(I_\ell, M/W)).$$

Consider now

$$\begin{tikzcd}
 & & H^1(G_{\mathbf{Q}_\ell}, M) \arrow[d] & \\
0 \arrow[r] & H^1_{ur}(G_{\mathbf{Q}_\ell}, M/W) \arrow[r] & H^1(G_{\mathbf{Q}_\ell}, M/W) \arrow[r] & H^1(I_\ell, M/W)
\end{tikzcd}$$

Then $L_\ell = ker(res \circ \theta)$.

The short exact sequence of $G_{\mathbf{Q}_\ell}$-modules

$$0 \to W \to M \to M/W \to 0$$

yields a long exact sequence of cohomology groups:

$$0 \to H^0(G_{\mathbf{Q}_\ell}, W) \to H^0(G_{\mathbf{Q}_\ell}, M) \to H^0(G_{\mathbf{Q}_\ell}, M/W) \to$$
$$H^1(G_{\mathbf{Q}_\ell}, W) \to H^1(G_{\mathbf{Q}_\ell}, M) \to Im(\theta) \to 0.$$

Since the alternating product of the orders is 1,

$$|Im(\theta)| = \frac{|H^0(G_{\mathbf{Q}_\ell}, W)||H^0(G_{\mathbf{Q}_\ell}, M/W)||H^1(G_{\mathbf{Q}_\ell}, M)|}{|H^0(G_{\mathbf{Q}_\ell}, M)||H^1(G_{\mathbf{Q}_\ell}, W)|}.$$

Next note that $|Im(res \circ \theta)| \geq |Im(\theta)|/|H^1_{ur}(G_{\mathbf{Q}_\ell}, M/W)| = |Im(\theta)|/|H^0(G_{\mathbf{Q}_\ell}, M/W)|$.

Putting this together,

$|L_\ell| = |H^1(G_{\mathbf{Q}_\ell}, M)|/|Im(res{\circ}\theta)| \leq |H^1(G_{\mathbf{Q}_\ell}, M)||H^0(G_{\mathbf{Q}_\ell}, M/W)|/|Im(\theta)|$ $= |H^0(G_{\mathbf{Q}_\ell}, M)||H^1(G_{\mathbf{Q}_\ell}, W)|/|H^0(G_{\mathbf{Q}_\ell}, W)|$.

Thus,

$$\frac{|L_\ell|}{|H^0(G_{\mathbf{Q}_\ell}, M)|} \leq \frac{|H^1(G_{\mathbf{Q}_\ell}, W)|}{|H^0(G_{\mathbf{Q}_\ell}, W)|} = |H^2(G_{\mathbf{Q}_\ell}, W)||F| = |H^0(G_{\mathbf{Q}_\ell}, W^*)||F|$$

(using respectively the local Euler characteristic, noting $v_\ell(|W|) = 1$, and local Tate duality).

To compute $|H^0(G_{\mathbf{Q}_\ell}, W^*)|$, let $c_\ell = (\psi_1(Fr_\ell)/\psi_2(Fr_\ell)) - 1$, which will turn out to be very important later! Let $\phi \in W^*$ be fixed by $G_{\mathbf{Q}_\ell}$, i.e. $\phi(gr) = g\phi(r)$ for all $g \in G_{\mathbf{Q}_\ell}, r \in F$. If $\alpha = \psi_1(Fr_\ell)/\psi_2(Fr_\ell)$, then by (†), $\phi(\alpha r) = \phi(r)$ ($\chi(g)$ cancels), i.e. $\phi(c_\ell r) = 0$, and so ϕ factors through $F/(c_\ell F)$. (This looks a little strange but later on we use the same calculation with $\bar{\rho}$ replaced by $\rho : G_{\mathbf{Q}} \to GL_2(\mathbf{Z}/\ell^n)$ and so F is replaced by R and $F/(c_\ell F)$ by $R/(c_\ell R)$.) In any case, $|H^0(G_{\mathbf{Q}_\ell}, W^*)| = |F/(c_\ell F)|$ and so

$$\frac{|L_\ell|}{|H^0(G_{\mathbf{Q}_\ell}, M)|} \leq |F||F/(c_\ell F)|.$$

The Big Picture. We obtained a formula, involving the orders of various Galois cohomology groups, for the tangential dimension r of $\Re_\Sigma$ (i.e. the smallest r such that $\Re_\Sigma$ is a quotient of $W(F)[[T_1, \ldots, T_r]]$). We also computed how this dimension changes under the operation of adding (or removing) a prime to/from Σ. We next intend to show that there is a universal lift of type Σ parametrizing lifts associated to modular forms,

$G_{\mathbf{Q}} \to GL_2(\mathcal{T}_\Sigma)$, and then show that the map $\Re_\Sigma \twoheadrightarrow \mathcal{T}_\Sigma$ this produces is an isomorphism of local rings. First, however, let us obtain a criterion for establishing such an isomorphism, a numerical criterion involving on the one hand r and on the other a certain invariant of $\mathcal{T}_\Sigma$.

8
Criteria for ring isomorphisms

Given the kind of $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(F)$ in which we are interested (i.e. absolutely irreducible and semistable) and a finite set Σ of rational primes, we have obtained a universal lift $\xi : G_{\mathbf{Q}} \to GL_2(\Re_\Sigma)$ of type Σ and we shall soon obtain a universal "modular" lift $\xi' : G_{\mathbf{Q}} \to GL_2(\mathcal{T}_\Sigma)$ of type Σ. By universality, we get a morphism $\phi_\Sigma : \Re_\Sigma \to \mathcal{T}_\Sigma$, which will be seen to be surjective. We want a numerical criterion that will suffice to show ϕ_Σ is an isomorphism.

By kind of $\bar{\rho}$, we also wish to include that it is associated to at least one cuspidal eigenform f. This means that f gives us a lift of $\bar{\rho}$ to characteristic zero, say $\rho_f : G_{\mathbf{Q}} \to GL_2(O)$, where O is the valuation ring of a finite extension of $\mathbf{Q}_\ell$ and is in $\mathcal{C}_F$. We shall see later that there exists f such that ρ_f is of type $\emptyset$, so of type Σ for any Σ. The universality of $\Re_\Sigma$ and $\mathcal{T}_\Sigma$ then yields a commutative triangle of morphisms:

$$\begin{tikzcd}
\Re_\Sigma \arrow[rr, "\phi"] \arrow[rr, "\Sigma"'] \arrow[dr] & & \mathcal{T}_\Sigma \arrow[dl] \\
& O &
\end{tikzcd}$$

Let $\mathcal{C}_O$ denote the subcategory of $\mathcal{C}_F$ consisting of local O-algebras (as noted earlier, every ring A in $\mathcal{C}_F$ is of the form $W(F)[[T_1,...,T_r]]/I$ and so is a local $W(F)$-algebra - we are asking that the map $W(F) \to A$ factors through O, or equivalently that A is a quotient of some $O[[T_1,...,T_r]]$). Inspired by our intended application, we shall consider the category $\mathcal{C}_O^*$, whose objects are pairs (A, π_A), where A is in $\mathcal{C}_O$ and $\pi_A : A \to O$ is a surjective morphism, and whose morphisms are morphisms $\phi : A \to B$ such that

$$\begin{tikzcd}
A \arrow[rr, "\phi"] \arrow[dr, "\pi_A"'] & & B \arrow[dl, "\pi_B"] \\
& O &
\end{tikzcd}$$

commutes.

Definition 8.1 *Associated to an object (A, π_A) of $\mathcal{C}_O^*$ are two invariants, namely the cotangent space $\Phi_A = (ker(\pi_A))/(ker(\pi_A))^2$ (a finitely generated O-module) and the congruence ideal $\eta_A = \pi_A(Ann_A ker(\pi_A))$ (an ideal of O).*

Call a ring A in $\mathcal{C}_O$ a complete intersection ring if A is free of finite rank as an O-module and is expressible as $O[[T_1,...,T_r]]/(f_1,...,f_r)$.

Example: Let f be the cuspidal eigenform associated to the elliptic curve denoted 57B in Cremona's book. This has level 57, weight 2, trivial Nebentypus, and all its coefficients in $\mathbf{Z}$. It therefore has an associated 3-adic representation $\rho_f : G_{\mathbf{Q}} \to$

$GL_2(\mathbf{Z}_3)$. Let $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(\mathbf{Z}/3)$ be the mod 3 representation it produces. One computes that $\bar{\rho}$ is surjective as follows.

The image of Fr_2 must have trace $a_2(f) = 1$ and determinant 2. The only such elements of $GL_2(\mathbf{Z}/3)$ have order 8. Next, the image of the local representation at 19 is given by Tate's theory of elliptic curves since 19 is a prime of good multiplicative reduction and we see that 3 divides its order. Thus 24 divides the order of the image of $\bar{\rho}$. One easily sees that the only subgroup of $GL_2(\mathbf{Z}/3)$ of order 24 is $SL_2(\mathbf{Z}/3)$ but the image of Fr_2 has nontrivial determinant. Thus the image is the whole of $GL_2(\mathbf{Z}/3)$ of order 48.

It follows that $\bar{\rho}$ is absolutely irreducible, Since $57B$ is a semistable curve, $\bar{\rho}$ is semistable. Our theory will apply to this example.

The theory of the next chapter allows us to compute that

$$\mathcal{T}_\emptyset = \{(x,y) \in \mathbf{Z}_3^2 | x \equiv y \pmod 3\} \cong \mathbf{Z}_3[[T]]/(T(T-3)).$$

Thus, $\mathcal{T}_\emptyset$ is a complete intersection ring, $\Phi_{\mathcal{T}_\emptyset} \cong \mathbf{Z}/3, \eta_{\mathcal{T}_\emptyset} = (3)$, and so $|\Phi_{\mathcal{T}_\emptyset}| = |\mathbf{Z}_3/\eta_{\mathcal{T}_\emptyset}|$.

Amazingly, these sorts of properties turn out to hold in general, and even more strikingly there is a simple numerical criterion that is sufficient to establish the kind of isomorphism we covet.

Theorem 8.2 *(Wiles, improved by Lenstra) Let $\phi : R \twoheadrightarrow T$ be a surjective morphism in $\mathcal{C}_O^*$. Assume that T is free of finite rank as an O-module and that $\eta_T \neq (0)$ (so $|O/\eta_T|$ is finite). Then the following are equivalent:*

(a) $|\Phi_R| \leq |O/\eta_T|$,

(b) $|\Phi_R| = |O/\eta_T|$,

(c) the map ϕ is an isomorphism of complete intersection

rings.

The theorem is established via the following lemma:

Lemma 8.3 *Suppose that A and B are in $\mathcal{C}_O^*$ and that there is a surjection $\phi : A \twoheadrightarrow B$.*

(1) ϕ induces a surjection $\tilde{\phi} : \Phi_A \twoheadrightarrow \Phi_B$ and so $|\Phi_A| \geq |\Phi_B|$. $\eta_A \subseteq \eta_B$.

(2) $|\Phi_A| \geq |O/\eta_A|$.

(3) Suppose B is a complete intersection ring. If $\tilde{\phi}$ is an isomorphism and Φ_A is finite, then ϕ is an isomorphism.

(4) Suppose A is a complete intersection ring. If $\eta_A = \eta_B \neq$ (0) and A and B are free, finite rank O-modules, then ϕ is an isomorphism.

(5) Suppose A is free and of finite rank as an O-module. Then there exists a complete intersection ring $\tilde{A}$ mapping onto A such that the induced map $\Phi_{\tilde{A}} \to \Phi_A$ is an isomorphism.

Proof: We show how the lemma implies that (a) implies (c). That (c) implies (b) will come out of the proof of the lemma later. That (b) implies (a) is trivial.

$$|O/\eta_T| \geq |\Phi_R| \geq |\Phi_T| \geq |O/\eta_T| \quad (*),$$

by what we are given, together with (1) and (2). Thus $|O/\eta_T| = |\Phi_T|$, whence

$$|O/\eta_T| = |\Phi_T| = |\Phi_{\tilde{T}}| \geq |O/\eta_{\tilde{T}}|,$$

by (5) and (2). But since by (1) $\eta_{\tilde{T}} \subseteq \eta_T$, it must be that $\eta_T = \eta_{\tilde{T}}$, which by (4) implies that the map $\tilde{T} \to T$ is an isomorphism and so T is a complete intersection ring.

Another consequence of $(*)$ is that $|\Phi_R| = |\Phi_T|$, and so by (3), $\phi : R \to T$ is an isomorphism. QED

Now we prove the lemma. For this we need first to know Nakayama's lemma and what Fitting ideals are.

Lemma 8.4 *(Nakayama's Lemma) Let A be a local ring and M a finitely generated A-module. If I is an ideal of A such that $IM = M$, then $M = 0$.*

Proof: Let $m_1, ..., m_n$ be a minimal generating set for M. Since $m_n \in M = IM$, we can write it $r_1 m_1 + ... + r_n m_n$ with $r_i \in I$. Then $(1 - a_n)m_n = r_1 m_1 + ... + r_{n-1} m_{n-1}$. We cannot have both a_n and $1 - a_n$ in the maximal ideal of A and so $1 - a_n$ is a unit, but then m_n is in the submodule of M generated by $m_1, ..., m_{n-1}$, a contradiction. QED

Corollary 8.5 *Suppose $\psi : M \to N$ is a homomorphism of finitely generated A-modules and I an ideal such that $M/IM \to N/IN$ is surjective. Then ψ is surjective.*

Proof: Apply Nakayama's lemma to the cokernel. QED

Note that $M/IM = M \otimes_A A/I$. In our context we shall apply the corollary by establishing that a map of A-modules is surjective after tensoring with O or F (which via π_A are quotients of A).

Definition 8.6 *Let R be in $\mathcal{C}_O$ and M be a finitely generated R-module. This means that there is an exact sequence*

$$0 \to M' \to R^n \to M \to 0$$

of R-modules (i.e. a presentation of M).

The Fitting ideal of M, $Fitt_R(M)$, is defined to be the ideal of R generated by $det(v_1, ..., v_n)$ as the $v_i \in R^n$ run through M'.

Exercise: (i) Show that this ideal is independent of the choice of presentation. [Hint: from two presentations $R^m \to M \to 0$, $R^n \to M \to 0$, form a presentation $R^{m+n} \to M \to 0$.]

(ii) Suppose $R = O$ and so without loss of generality

$$M = O^r \oplus (O/\lambda^{n_1}) \oplus \ldots \oplus (O/\lambda^{n_k}),$$

where λ is a uniformizer of O. Show that $Fitt_O(M) = (0)$ if $r > 0$ and $= (\lambda^{n_1 + \ldots n_k})$ if $r = 0$.

Deduce that $|M| = |O/Fitt_O(M)|$.

Proof: (1) The composition $ker(\pi_A) \twoheadrightarrow ker(\pi_B) \twoheadrightarrow \Phi_B$ factors through Φ_A. The mere existence of a map $Ann_A ker(\pi_A) \to Ann_B ker(\pi_B)$ implies that $\eta_A \subseteq \eta_B$.

(2) We note that $Fitt_R(M) \subseteq Ann_R(M)$. This follows because if we take a presentation

$$0 \to M' \to R^n \to M \to 0$$

and let $x_1, ..., x_n \in M$ be the images of the standard generators of R^n and let $d = det(c_{ij})$ where $(c_{ij}) = (v_1, \ldots, v_n)$ with each $v_i \in M'$, then $\sum_j c_{ij} x_j = 0$ and so $dI_n = (d_{ij})(c_{ij})$ (where (d_{ij}) is the adjoint-transpose of (c_{ij})) multiplies the vector with entries x_i to 0, whence $dx_i = 0$ for $1 \le i \le n$, and so since the x_i generate M, $d \in Ann_R(M)$.

Secondly, note that $\pi_A(Fitt_A(M)) = Fitt_O(M \otimes_A O)$ (where π_A makes O into an A-algebra), since as an O-module $M \otimes_A O = M/(ker(\pi_A)M)$ is defined by the same relations as those defining M as an A-module).

In particular, $ker(\pi_A) \otimes_A O = \Phi_A$, and so

$$Fitt_O(\Phi_A) = \pi_A(Fitt_A(ker(\pi_A)) \subseteq \pi_A(Ann_A ker(\pi_A)) = \eta_A,$$

whence $|\Phi_A| = |O/Fitt_O(\Phi_A)| \ge |O/\eta_A|$, the first equality coming out of the exercise above.

[Note that the "(a) $\iff$ (b)" part of the main theorem follows since if R surjects onto T, then always $|\Phi_R| \geq |\Phi_T| \geq |O/\eta_T|$.]

(3) Consider $U = O[[T_1, ..., T_r]]$ as in $\mathcal{C}_O^*$ by taking π_U to be the map sending each $T_i \mapsto 0$. Since B is a complete intersection ring, there exists a local homomorphism $\nu_B : U \to B$ with kernel $(f_1, ..., f_r)$. Letting $b_i = \nu_B(T_i)$, then $b_i \in ker(\pi_B)$. Since $\tilde{\phi} : \Phi_A \to \Phi_B$ is an isomorphism, there exists $a_i \in ker(\pi_A)$ mapping to b_i and the images of a_i in Φ_A generate Φ_A.

Now define $\nu_A : U \to A$ by sending $T_i \mapsto a_i$. This gives a surjection $\bar{\nu}_A : \Phi_U \to \Phi_A$ and so by Nakayama ν_A is surjective. We next establish that $ker(\nu_B) \subseteq ker(\nu_A)$ (and so are actually equal).

Since $\Phi_U \cong O^r$ and Φ_A is finite, $ker(\bar{\nu}_A)$ has r generators, say $\bar{g}_1, \ldots, \bar{g}_r$. Pick $g_1, \ldots, g_r \in ker(\nu_A)$ mapping to $\bar{g}_1, \ldots, \bar{g}_r$. Since $ker(\nu_A) \subseteq ker(\nu_B)$, $(g_1, ..., g_r) = (f_1, ..., f_r)M$ for some $M \in M_r(U)$. Let $\bar{M} = M \pmod{(T_1, ..., T_r)}$, i.e. the matrix of constant terms. Then $(\bar{g}_1, \ldots, \bar{g}_r) = (\bar{f}_1, \ldots, \bar{g}_r)\bar{M}$. Since these generate the same rank r submodule of finite index in Φ_U, $det(\bar{M}) \in O^\times$. Thus M is invertible and we are done.

Finally, this means that $\nu_A \nu_B^{-1}$ is well-defined. It is an inverse to ϕ and so ϕ is an isomorphism.

(4) We first claim that $ker(\pi_A) \cap Ann_A ker(\pi_A) = 0$. This is proven as follows. Suppose $x \in \eta_A, x \neq 0$. Say $x = \pi_A(x')$, where $x' \in Ann_A ker(\pi_A)$. If $a \in ker(\pi_A) \cap Ann_A ker(\pi_A)$, then since $x - x' \in ker(\pi_A)$, $0 = a(x - x') = ax$ (since $a \in ker(\pi_A)$ and $x' \in Ann_A ker(\pi_A)$). Thus a is O-torsion, so $a = 0$.

Thus the restriction $\pi_A : Ann_A ker(\pi_A) \to \eta_A$ is injective. The definition of η_A makes it surjective. Since $\eta_A = \eta_B$, this isomorphism translates into the restriction of $\phi : Ann_A ker(\pi_A) \to$

$Ann_B ker(\pi_B)$ being an isomorphism.

So we have an exact sequence $0 \to ker(\phi) \oplus Ann_A ker(\pi_A) \to A$, with the cokernel $A/(ker(\phi) \oplus Ann_A ker(\pi_A)) \cong B/(\phi(Ann_A ker(\pi_A)) \cong B/(Ann_B ker(\pi_B)) \hookrightarrow End_O ker(\pi_B)$. Thus the cokernel is torsionfree, giving us splitting over O. If we now define $A^\wedge = \hom_O(A, O)$, then one can see explicitly that $A^\wedge \cong A$ as A-modules (this says that complete intersection rings are *Gorenstein*). We therefore get a dual exact sequence:

$$A \to (ker(\phi))^\wedge \oplus (Ann_A ker(\pi_A))^\wedge \to 0.$$

Tensoring with F (the residue ring of A) gives $1 = dim_F(A \otimes_A F)$ and $(Ann_A ker(\pi_A))^\wedge \otimes_A F \neq 0$ (since $\eta_A \neq 0$). Thus $(ker(\phi))^\wedge \otimes_A F = 0$. By Nakayama's lemma and dualizing, $ker(\phi) = 0$, and we are done.

(5) As seen in (2), we can write A as a quotient of $U = O[[T_1, ..., T_r]]$, where the T_i map to elements of $ker(\pi_A)$. The idea is to choose $\bar{f}_1, ..., \bar{f}_r$ generating the kernel of the induced $\Phi_U (= O^r) \to \Phi_A$ and then to lift these to U and set $\tilde{A} = U/(f_1, ..., f_r)$, a complete intersection ring if we ensure that $\tilde{A}$ is finitely generated.

Let $a_1, ..., a_r$ be O-module generators of $ker(\pi_A)$. Let $V = O[T_1, ..., T_r]$. Define $\phi : V \to A$ by $T_j \mapsto a_j$. Then ϕ is surjective. Pick $f_1, ..., f_r \in ker(\phi)$ and let m be the maximal degree occurring. Since $a_i^2 \in ker(\pi_A)$, $a_i^2 = h_i(a_1, ..., a_r)$ for some linear polynomial h_i. Replace f_i by $f_i + T_i^m h_i - T_i^{m+2}$. Then $V/(f_1, ..., f_r)$ is finitely generated as an O-module. Complete at $(\lambda, T_1, ..., T_r)$ to get $\tilde{A} = U/(f_1, ..., f_r)$, a finitely generated O-module. Note that $\tilde{A} \to A$ induces an isomorphism on the cotangent spaces since the linear terms of the f_i generate the kernel of the induced map $\Phi_U \to \Phi_A$.

QED

There is one problem in applying the Wiles-Lenstra criterion to our set-up, namely that $\Re_\Sigma$ and $\mathcal{T}_\Sigma$ are certainly $W(F)$-algebras but may or may not be O-modules. We can fix this by setting $R = R_\Sigma \otimes_{W(F)} O$ and $T = \mathcal{T}_\Sigma \otimes_{W(F)} O$, and checking the following exercise.

Exercise:

(a) $\Re_\Sigma \to \mathcal{T}_\Sigma$ is an isomorphism if and only if the induced map $R \to T$ is an isomorphism.

(b) T_Σ is a complete intersection ring if and only if T is one.

(c) R and T are the universal deformation rings (for type Σ and type Σ modular lifts respectively) if we restrict Mazur's functor to $\mathcal{C}_O$.

With these rings R and T, we now need to interpret Φ_R and η_T with the aim of proving $|\Phi_R| \leq |O/\eta_T|$.

Lemma 8.7 *Let E be the fraction field of O. There is a natural bijection from*

$$Hom_O(\Phi_R, E/O) \to H^1_\Sigma(G_{\mathbf{Q}}, \mathrm{Ad}^0(\rho_f) \otimes E/O).$$

Thus, $|Phi_R| = |H^1_\Sigma(G_{\mathbf{Q}}, \mathrm{Ad}^0(\rho_f) \otimes E/O)|$.

Proof: Recall how at the start of chapter 7 we showed how given $\bar{\rho} : G \to GL_n(F)$ a lift $\sigma : G \to GL_n(F[\epsilon])$ yielded a 1-cocycle $G \to ker(GL_n(F[\epsilon]) \to GL_n(F))$ by considering the difference between σ and the trivial lift of $\bar{\rho}$. Likewise, here we have a trivial lift of $\rho_f : G_{\mathbf{Q}} \to GL_2(O)$ to $R/\wp^2$, where $\wp = ker(R \to O)$, since R (and so $R/\wp^2$) is an O-algebra. Using this, each lift of ρ_f to $R/\wp^2$ gives a 1-cocycle $G_{\mathbf{Q}} \to ker(GL_2(R/\wp^2) \to GL_2(O)) = M_2(\Phi_R)$. Strictly equivalent lifts differ by a 1-coboundary and the process can be reversed.

Next, note that $f \in Hom(\Phi_R, O/\lambda^n)$ together with this defines a class in $H^1_\Sigma(G_{\mathbf{Q}}, \mathrm{Ad}^0(\rho_f) \otimes O/\lambda^n)$, since we are only considering lifts of type Σ. Taking the union (direct limit) over all n yields the lemma. Another way of seeing this is to consider lifts of ρ_f to $O[\epsilon]/(\lambda^n\epsilon, \epsilon^2)$. Every such lift yields a map $R \to O[\epsilon]/(\lambda^n\epsilon, \epsilon^2)$. This restricts to a homomorphism $\wp \to (O/\lambda^n)\epsilon$ such that $\wp^2$ maps to 0, and so this factors through Φ_R. QED

We can now use the Wiles-Lenstra criterion to reduce proving that $\Re_\Sigma \to \mathcal{T}_\Sigma$ is always an isomorphism to the case $\Sigma = \emptyset$. Let $\Sigma' = \Sigma \cup \{p\}, R' = \Re_{\Sigma'} \otimes_{W(F)} O, T' = \mathcal{T}_{\Sigma'} \otimes_{W(F)} O$. We have the following commutative diagram, reflecting the facts that the larger rings parametrize larger sets:

$$\begin{tikzcd}
R' \arrow[r] \arrow[d] & T' \arrow[d] \\
R \arrow[r] & T
\end{tikzcd}$$

Theorem 8.8 *There exists $c_p \in O$ satisfying*

$$|\Phi_{R'}|/|\Phi_R| \leq |O/c_p|, \quad \eta_{T'} \subseteq c_p\eta_T.$$

We shall specify c_p shortly. The amazing thing is that the same c_p arises on either side. This is what makes our reduction work.

Corollary 8.9 *Suppose $R \twoheadrightarrow T$ is an isomorphism. Then so is $R' \twoheadrightarrow T'$. Applying this one prime at a time, we see that if $\Re_\emptyset \twoheadrightarrow \mathcal{T}_\emptyset$ is an isomorphism, then $\Re_\Sigma \twoheadrightarrow \mathcal{T}_\Sigma$ is an isomorphism for every finite set Σ of primes.*

Proof: By the theorem, $|\Phi_{R'}| \leq |\Phi_R||O/c_p|$, which $= |O/\eta_T||O/c_p|$

by Wiles-Lenstra since $R \to T$ is an isomorphism. Applying the last theorem again, this $\leq |O/\eta_{T'}|$. Applying Wiles-Lenstra again gives that $R' \to T'$ is an isomorphism. QED

The big question is what c_p works. If $p \neq \ell$, then following theorem 7.15 we have that

$$|\Phi_{R'}|/|\Phi_R| \leq |H^0(G_{\mathbf{Q}_p}, M^*)|,$$

where $M = \mathrm{Ad}^0(\rho_f) \otimes E/O$. This H^0 is just the fixed points of M^* under the action of $G_{\mathbf{Q}_p}$ and so is readily computed.

Lemma 8.10 *Suppose $p \neq \ell$. $|H^0(G_{\mathbf{Q}_p}, M^*)| = |O/c_p|$ where $c_p = (p-1)(a_p^2 - (p+1)^2)$ if p is unramified in $\bar{\rho}$ (and so ρ_f), and $= p^2 - 1$ otherwise.*

Proof: In the case that p is unramified, I_p acts trivially and so the action of $G_{\mathbf{Q}_p}$ factors through $< Fr_p >$. Thus we need the order of the submodule of M^* fixed under Fr_p. The first thing to note is that the representation $\mathrm{Ad}^0(\rho_f) = Sym^2 \otimes det^{-1}$ and so we just need the fixed points of Sym^2. The eigenvalues of Fr_p are $\alpha_p^2, \alpha_p\beta_p, \beta_p^2$ as given below and so of $1 - Fr_p$ are $1-\alpha_p^2, 1-\alpha_p\beta_p, 1-\beta_p^2$. Thus the determinant of the action is $(1-\alpha_p^2)(1-\alpha_p\beta_p)(1-\beta_p^2) = c_p$ above.

The ramified at p case is similar. QED

Remarks: Note that $c_p \neq 0$. This follows from the Petersson-Ramanujan bound $|a_p| \leq 2\sqrt{p}$.

This will be our choice of c_p for $p \neq \ell$. To give some perspective of how Wiles' desired inequality actually fits into the big picture of number theory, relating orders of Selmer groups to special values of L-functions, we next define the L-function of the symmetric square of ρ_f.

Definition 8.11 *Suppose $f = \sum a_n q^n$ is a cuspidal eigenform* *of weight* 2 *and trivial Nebentypus. Let* α_p, β_p *be the eigenvalues of Frobenius, so they satisfy* $\alpha_p + \beta_p = a_p(f)$ *and* $\alpha_p\beta_p = p$*. The* p*th Euler factor of* $L(s, \rho_f)$ *is* $1/((1-\alpha_p p^{-s})(1-\beta_p p^{-s})) = 1/(1 - a_p p^{-s} + pp^{-2s})$*. The* p*th Euler facor of* $L(s, Sym^2(\rho_f))$ *is* $1/((1-\alpha^2 p^{-s})(1-\alpha_p\beta_p p^{-s})(1-\beta^2 p^{-s}))$*.*

The amazing fact is that up to a power of p, c_p is the pth Euler factor of $L(2, Sym^2(\rho_f))$. In this way we can restate the desired inequality in terms of the order of a Selmer group being bounded by a special value of an L-function, and we have a case of the Bloch-Kato conjecture, a vast generalization of the Birch and Swinnerton-Dyer conjecture. Unfortunately, this observation led Wiles off on a wild goose chase for several years, seeking to prove this case of Bloch-Kato by generalizing ideas of Flach to construct geometric Euler systems, similar to those that had been successful in proving cases of the Birch and Swinnerton-Dyer conjecture earlier. We shall return to this observation once we have introduced $\mathcal{T}_\Sigma$.

Lemma 8.12 *If* $p = \ell$*, then if* $\bar\rho$ *is not both good and ordinary, then* $|H^1_{ss}|/|H^1_f| = 1$ *and so we set* $c_\ell = 1$*. If* $\bar\rho$ *is both good and ordinary, then we set* $c_\ell = a_\ell^2 - (\ell+1)^2$*. Up to units, this is the same as* $\alpha_\ell^2 - 1$*, where* α_ℓ *is the unit root of* $X^2 - a_\ell X + \ell$ *(there is one such by the ordinariness of* $\bar\rho$ *at* ℓ*.*

Proof: The factor in Wiles' formula for the order of the Selmer group goes up by $|H^1_{ss}|/|H^1_f|$ in going from Σ to Σ'. If $\bar\rho$ is not good, then in definition 7.6 H^1_f was taken to be H^1_{ss}. If $\bar\rho$ is not ordinary, then its lifts are not either and so a lift is semistable if and only if it is good.

In the good, ordinary case, we already calculated that, if $M = \mathrm{Ad}^0(\rho_f) \otimes O/\lambda^n$, then $|H^1_f(G_{\mathbf{Q}_\ell}, M)| = |H^0(G_{\mathbf{Q}_\ell}, M)||O/\lambda^n|$, whereas $|H^1_{ss}(G_{\mathbf{Q}_\ell}, M)| \leq |H^0(G_{\mathbf{Q}_\ell}, M)||O/\lambda^n||O/(\lambda^n, c_\ell)|$, where $c_\ell = (\psi_1/\psi_2)(Fr_\ell) - 1$. Since $\psi_1 = \psi_2^{-1}$ and $\psi_2(Fr_\ell)$ is the unit root of $X^2 - a_\ell X + \ell$, we get $c_\ell = \alpha_\ell^2 - 1$. QED

Once we describe the construction of $\mathcal{T}_\Sigma$ and hence compute η_T, we shall be able to show the other half - that $\eta_{T'} \subseteq c_p \eta_T$. This then reduces us to having to prove the isomorphism in the case of $\Sigma = \emptyset$. For this we prove the following. In applications, R_n and T_n will be rings obtained by using sets Σ containing primes of a specific form, for which we can control the structure of $\Re_\Sigma$ and $\mathcal{T}_\Sigma$. The proof of this result, while considerably easier than the earlier proposed approach via “Flach systems”, is still quite involved and forms the content of the companion paper by Taylor and Wiles [34]. It is where the gap that remained for 16 months, is fixed.

Theorem 8.13 *(Wiles, improved by Faltings) Let $\phi : R \twoheadrightarrow T$ be a surjection in $\mathcal{C}^*_O$ with T finite and free over O. For any $r \geq 0$ regard R and T as $O[[S_1, ..., S_r]]$-modules by letting the S_i act trivially. Let $\omega_n(T)$ denote $(1+T)^{p^n} - 1$.*

Suppose there exists an integer $r \geq 0$ such that for every $n \geq 1$ there is a commutative diagram of surjective homomorphisms of local $O[[S_1, ..., S_r]]$-algebras

$$\begin{tikzcd}
R_n \arrow[r] \arrow[d] & T_n \arrow[d] \\
R \arrow[r, "\phi"] & T
\end{tikzcd}$$

that has the following properties

(i) the induced maps $R_n/(S_1, ..., S_r)R_n \to R$ and $T_n/(S_1, ..., S_r)T_n \to T$ are isomorphisms;

(ii) the O-algebras R_n can be generated by r elements;

(iii) the rings $T_n/(\omega_n(S_1), ..., \omega_n(S_r))T_n$ *are finite free* $O[[S_1, ..., S_r]]/(\omega_n(S_1), ..$ *algebras.*

Then $\phi : R \to T$ *is an isomorphism of complete intersection rings.*

9
The universal modular lift

Fix a continuous homomorphism $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(F)$, where F is a finite field of characteristic $\ell \geq 3$. We assume as usual that $\bar{\rho}$ is absolutely irreducible, semistable (and so in particularΦ odd), and modular (i.e. there is a cuspidal eigenform $f = \sum a_n q^n$ such that $tr\bar{\rho}(Fr_p) = a_p$ for all but finitely many primes p). Letting $K = \mathbf{Q}(\sqrt{\pm\ell})$, where we pick the positive sign if $\ell \equiv 1 \pmod 4$ and the negative sign otherwise, we shall also want to assume that the restriction of $\bar{\rho}$ to G_K is absolutely irreducible. In fact our assumptions already imply this if $\ell \geq 5$ but not necessarily in the critical case of interest to us, $\ell = 3$.

Fix a finite set Σ of primes. We wish to obtain a lift of type Σ that parametrizes all lifts of type Σ associated to cuspidal eigenforms.

Recall that $S_2(N)$ denotes the $\mathbf{C}$-vector space of cusp forms of weight 2, level N, and trivial Nebentypus. Let $\mathcal{T}' = \mathcal{T}'(N)$ denote the ring of endomorphisms of this space generated by the Hecke operators T_n for n prime to ℓN. Recall what has been shown already.

Theorem 9.1 *Let* $\mathbf{m}$ *be a maximal ideal of* $\mathcal{T}'$ *and suppose* $\mathcal{T}'/\mathbf{m}$ *has characteristic* ℓ.

(a) There exists a continuous semisimple homomorphism $\tilde{\rho_{\mathbf{m}}} : G_{\mathbf{Q}} \to GL_2(\mathcal{T}'/\mathbf{m})$, *unramified outside* ℓN, *such that* $tr\tilde{\rho_{\mathbf{m}}}(Fr_p) = T_p$ *and* $det\tilde{\rho_{\mathbf{m}}}(Fr_p) = p$ *for all* $p \not| \ell N$.

(b) Suppose $\tilde{\rho_{\mathbf{m}}}$ *is absolutely irreducible. Let* $\mathcal{T}'_{\mathbf{m}}$ *be the* $\mathbf{m}$*-adic completion of* $\mathcal{T}'$. *There exists a continuous homomorphism* $\rho_{\mathbf{m}} : G_{\mathbf{Q}} \to GL_2(\mathcal{T}'_{\mathbf{m}})$, *unramified outside* ℓN, *such that* $tr\rho_{\mathbf{m}}(Fr_p) = T_p$ *and* $det\rho_{\mathbf{m}}(Fr_p) = p$ *for all* $p \not| \ell N$.

Proof: Most of this was established in theorem 4.28, where we obtained a free $\mathcal{T} \otimes_{\mathbf{Z}} \mathbf{Q}_\ell$-module of rank 2 on which $G_{\mathbf{Q}}$ acts as above. The injection $\mathcal{T}' \hookrightarrow \mathcal{T}$ yields an injection $\mathcal{T}'_{\mathbf{m}} \hookrightarrow \mathcal{T} \otimes_{\mathbf{Z}} \mathbf{Q}_\ell$. Since the traces of Frobenius land in $\mathcal{T}'_{\mathbf{m}}$, by Chebotarev's density theorem (that the Frobenius elements generate $G_{\mathbf{Q},S}$ for any finite S) the trace of any element lies in $\mathcal{T}'_{\mathbf{m}}$. This together with the absolute irreducibility of $\tilde{\rho_{\mathbf{m}}}$ gives, by a result of Carayol, that there is a representation into $GL_2(\mathcal{T}'_{\mathbf{m}})$ with the same trace as our given representation into $GL_2(\mathcal{T} \otimes_{\mathbf{Z}} \mathbf{Q}_\ell)$. QED

Now let $\bar{N} = \bar{N}(\bar{\rho})$ be the product of the primes at which $\bar{\rho}$ fails to be good. Since $\bar{\rho}$ is semistable, $N(\bar{\rho})$ is squarefree and so $\bar{N} = N(\bar{\rho})$ if $\bar{\rho}$ is good at ℓ, and $\ell N(\bar{\rho})$ otherwise.

Theorem 9.2 *There exists a unique ring homomorphism* $a : \mathcal{T}'(\bar{N}) \to F$ *such that* $a(T_p) = tr\bar{\rho}(Fr_p)$ *for all* $p \not| \ell \bar{N}$. *Let* $m = kera$, *a maximal ideal of* $\mathcal{T}'(\bar{N})$. *Then* $\bar{\rho}$ *is isomorphic to* $\tilde{\rho}_{\mathbf{m}} : G_{\mathbf{Q}} \to GL_2(\mathcal{T}'(\bar{N})/\mathbf{m}) \to GL_2(F)$.

Proof: This is simply a restatement of the work of Ribet and others showing that if $\bar{\rho}$ is modular (as we have assumed), then it is associated to a cuspidal eigenform f of weight, level, and Nebentypus predicted by Serre's invariants, in this case weight 2, level $\bar{N}$, and trivial Nebentypus. Now define a as the usual eigencharacter, $T(f) = a(T)f$ for $T \in \mathcal{T}'(\bar{N})$. QED

Next we introduce Σ. Let $N_\Sigma = N_\Sigma(\bar{\rho}) = \prod p^{n_p}$, where n_p is the exponent of p in $\bar{N}$ if $p \notin \Sigma$, $= 2$ if $p \in \Sigma, p \neq \ell$, and $= 1$ if $p \in \Sigma, p = \ell$. In particular, $N_\emptyset = \bar{N}$. Since $\bar{N} | N_\Sigma$, there is a map $r : \mathcal{T}'(N_\Sigma) \to \mathcal{T}'(\bar{N})$. Set $\mathbf{m}_\Sigma = r^{-1}(\mathbf{m})$, a maximal ideal of $\mathcal{T}'(N_\Sigma)$.

Theorem 9.3 *Let ρ be a lift of $\bar{\rho}$ to a ring A in $\mathcal{C}_F$. The following are equivalent:*

(a) ρ is unramified outside ℓN_Σ and there exists a ring homomorphism $\alpha : \mathcal{T}'(N_\Sigma) \to A$ such that $tr\rho(Fr_p) = \alpha(T_p)$ for all $p \not| \ell N_\Sigma$.

(b) There exists a ring homomorphism $\hat{\alpha} : \mathcal{T}'(N_\Sigma)_{\mathbf{m}_\Sigma} \to A$ such that ρ is isomorphic to the representation obtained by composition of $\rho_{\mathbf{m}_\Sigma}$ with $\hat{\alpha}$.

When these hold, (a) determines α uniquely, $\hat{\alpha}$ extends α continuously, and ρ is a lift of $\bar{\rho}$ of type Σ.

Proof: (b) implies (a): Composition of the completion map $\mathcal{T}'(N_\Sigma) \to \mathcal{T}'(N_\Sigma)_{\mathbf{m}_\Sigma}$ with $\hat{\alpha}$ yields α (whose traces of Frobenius are as stated since this holds for $\rho_{\mathbf{m}_\Sigma}$).

(a) implies (b): Since T_p maps to $tr\bar{\rho}(Fr_p)$ (p not in Σ and not dividing $\ell N(\bar{\rho})$) either way around, the following diagram commutes:

$$\begin{tikzcd}
\mathcal{T}'(N_\Sigma) \arrow[r, "\alpha"] \arrow[d] & A \arrow[d] \\
\mathcal{T}'(\bar{N}) \arrow[r, "a"] & F
\end{tikzcd}$$

It follows that $\alpha(\mathbf{m}_\Sigma)$ lands in the maximal ideal of A, which is exactly what we need for α to have a continuous extension to $\hat{\alpha}$.

The tricky thing here is to establish that ρ is of type Σ. This is controlled by the form of N_Σ. Namely, $J_0(N_\Sigma)$ has good reduction at all primes not dividing N_Σ and semistable reduction at all primes p such that $p^2 \not| N_\Sigma$. Our form of N_Σ then implies that the reduction at ℓ is semistable and that the reduction at any prime $\notin \Sigma$ is the same as that for $\bar{\rho}$. The final piece needed is an analogue of Tate's elliptic curves for semistable abelian varieties (due to Grothendieck, LNM 288) that translates these reduction properties over to the form of the associated local Galois representations, just as we did in chapter 5. QED

Definition 9.4 *Call a lift of $\bar{\rho}$ satisfying the equivalent conditions of the previous theorem a modular lift of $\bar{\rho}$ of type Σ.*

Now we are in a position to produce a universal such lift. Let F_0 be the subfield of F generated by $\{tr\bar{\rho}(g)|g \in G_{\mathbf{Q}}\}$, which is by Chebotarev the same as the subfield generated by the traces of Frobenius. Thus $\mathcal{T}'(N_\Sigma)/\mathbf{m}_\Sigma \cong F_0$. Set $\mathcal{T}_\Sigma = \mathcal{T}'(N_\Sigma)_{\mathbf{m}_\Sigma} \otimes_{W(F_0)} W(F)$.

Then $\mathcal{T}_\Sigma$ is an object in $\mathcal{C}_F$ and the map $\mathcal{T}'(N_\Sigma)_{\mathbf{m}_\Sigma} \to \mathcal{T}_\Sigma$ composed with $\rho_{\mathbf{m}_\Sigma} : G_{\mathbf{Q}} \to GL_2(\mathcal{T}'(N_\Sigma)_{\mathbf{m}_\Sigma})$ yields a lift $\rho_\Sigma^{mod} : G_{\mathbf{Q}} \to GL_2(\mathcal{T}_\Sigma)$ of $\bar{\rho}$ which is modular of type Σ.

Then the previous theorem translates to the following (which says that ρ_Σ^{mod} is the universal lift of $\bar{\rho}$ of type Σ):

Theorem 9.5 *Given a modular lift ρ of $\bar{\rho}$ of type Σ to a ring A in $\mathcal{C}_F$, there exists a unique morphism $\phi : \mathcal{T}_\Sigma \to A$ such that ρ_Σ^{mod} composed with ϕ is strictly equivalent to ρ.*

Remarks: (1) If $\Sigma \subseteq \Sigma'$, then by universality we get a unique morphism $\mathcal{T}_{\Sigma'} \to \mathcal{T}_\Sigma$ sending $\rho_{\Sigma'}^{mod}$ to ρ_Σ^{mod}. This is surjective since $\mathcal{T}_\Sigma$ is generated by the traces of the representation.

(2) Also by universality, there exists $\phi_\Sigma : \Re_\Sigma \to \mathcal{T}_\Sigma$ sending the universal lift of type Σ to ρ_Σ^{mod}. By the same reason as in (1), ϕ_Σ is surjective.

(3) Since $\mathcal{T}'(N_\Sigma)$ is reduced (i.e. no nilpotents other than 0) and a free $\mathbf{Z}$-module of finite rank (recall this follows since it is a subring of $EndS_2(N, \mathbf{Z})$ by the q-expansion principle), $\mathcal{T}_\Sigma$ is reduced and a free $W(F)$-module of finite rank.

We can even give an explicit construction of $\mathcal{T}_\Sigma$. Given $\bar{\rho}$, consider the set of newforms f such that $\rho_f : G_{\mathbf{Q}} \to GL_2(O_f)$ is equivalent to a lift of $\bar{\rho}$ of type Σ. This is nonempty since by Ribet there is such a cuspidal eigenform giving a lift of type $\emptyset$, and so necessarily of type Σ. Inside $\prod O_f$ consider for each prime p not in Σ and not dividing $\ell N(\bar{\rho})$ the element $(a_p(f))$. Then $\mathcal{T}_\Sigma$ is the $W(F)$-algebra generated by all such elements.

Example: Continuing our example where $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(\mathbf{Z}/3)$ is associated to the action on the 3-division points of the elliptic curve E denoted $57B$ in Cremona's tables, $y^2 + y = x^3 - 2x - 1$. Let f_E denote the associated modular form. We need to consider what other newforms of level dividing 57 and weight 2 produce the same $\bar{\rho}$, which is the same as being congruent (mod 3) to f_E. As noted in Cremona's tables, there is precisely one other elliptic curve of conductor dividing 57 that produces such a newform, namely $57C$. Since the forms are congruent (mod 3) but not (mod 9), we get that

$$\mathcal{T}_\emptyset \cong \{(x,y) \in \mathbf{Z}_3^2 | x \equiv y \pmod 3\} \cong \mathbf{Z}_3[[T]]/(T(T-3)).$$

In particular, this is a complete intersection ring and $\eta_{\mathcal{T}_\emptyset} = 3\mathbf{Z}_3$.

If, in a general situation, there are exactly two newforms f and g (of allowable weight and level) producing the same $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(\mathbf{Z}/\ell)$ and they are congruent $\pmod{\ell^n}$ but not $\pmod{\ell^{n+1}}$, then $\mathcal{T}_\emptyset \cong \mathbf{Z}_\ell[[T]]/(T(T-\ell^n))$ yielding $\eta_{\mathcal{T}_\emptyset} = \ell^n\mathbf{Z}_\ell$. In general, $\eta_{\mathcal{T}_\Sigma}$ measures congruences between the corresponding newforms, and this is why it is called the congruence ideal.

Besides the properties obtained so far for $\mathcal{T}_\Sigma$ (see remark (3) above), we need to establish three further properties of $\mathcal{T}_\Sigma$. The first one we needed in reducing to the minimal case. Let $T = \mathcal{T}_\Sigma$ and $T' = \mathcal{T}_{\Sigma'}$, where $\Sigma' = \Sigma \cup \{p\}$.

Theorem 9.6 *With the choice of $c_p \in O$ from the previous chapter, $\eta_{T'} = c_p \eta_T$.*

Proof: Given $\rho_\Sigma^{mod} : G_{\mathbf{Q}} \to GL_2(\mathcal{T}_\Sigma)$, let M_Σ denote the corresponding $G_{\mathbf{Q}}$-module, a free rank 2 $\mathcal{T}_\Sigma$-module. It comes with an alternating pairing (the Weil pairing) $M_\Sigma \times M_\Sigma \to O$, which we shall denote by $(x,y) \mapsto < x, y >_\Sigma$. Let $\wp_\Sigma = ker(\mathcal{T}_\Sigma \to O)$ and $L_\Sigma = M_\Sigma[\wp_\Sigma]$ (i.e. the elements of M_Σ annihilated by $\wp_\Sigma$), a free $T_\Sigma/\wp_\Sigma = O$-module of rank 2.

Claim: If $\{x, y\}$ is a basis for L_Σ, then $\eta_T =< x, y >_\Sigma$.

To compare η_T and $\eta_{T'}$, we need to construct a homomorphism $\beta : M_{\Sigma'} \to M_\Sigma$. Consider for now the most general case $p \neq \ell, p \not| N(\bar{\rho})$. Then $N_{\Sigma'} = N_\Sigma p^2$. Let $X = X_0(N_\Sigma), X' = X_0(N_{\Sigma'})$, and J, J" be the respective Jacobians. The map on the upper half-plane sending $\tau \mapsto (\tau, p\tau, p^2\tau)$ induces "degeneracy maps" $X' \to X^3$. By Albanese functoriality, this yields a map $J' \to J^3$.

(In the cases $p \neq \ell, p|N(\bar{\rho})$ or $p = \ell, \bar{\rho}$ good and ordinary, we do the same with just $\tau \mapsto (\tau, p\tau)$. For $p = \ell, \bar{\rho}$ otherwise, we use just $\tau \mapsto \tau$.)

This then induces a homomorphism $T_\ell(J') \otimes_{\mathbf{Z}_\ell} O \to (T_\ell(J) \otimes_{\mathbf{Z}_\ell} O)^3$. By picking the right map $(M_\Sigma)^3 \to M_\Sigma$, namely $(1, -p^{-1}T_p, p^{-1})$, we induce a map $\beta : M_{\Sigma'} \to M_\Sigma$. Let $\beta' : M_\Sigma \to M_{\Sigma'}$ be the adjoint of β, i.e. satisfy

$$< x, \beta y >_\Sigma = < \beta' x, y >_{\Sigma'}$$

Wiles computes $\beta\beta'$ by looking at the effect on cuspforms. He finds that $\beta\beta' = -p^{-2}(p-1)(T_p^2 - (p+1)^2)$, which up to a unit is c_p. In every other case, the same thing holds.

Now, assuming that if $\{x, y\}$ is a basis of M_Σ, then $\{\beta(x), \beta(y)\}$ is a basis of $M_{\Sigma'}$ (which is a tricky problem in cohomology of modular curves to show that β is surjective), then we have

$$\eta_{\Sigma'} = (< \beta'(x), \beta'(y) >_{\Sigma'}) = (< x, \beta\beta'(y) >)_\Sigma = c_p(< x, y >_\Sigma) = c_p\eta_\Sigma$$

QED

10
The minimal case

Our aim then is first to show that the homomorphism $\Re_\emptyset \to \mathcal{T}_\emptyset$ is an isomorphism. This turned out to be the hardest part of the proof and the part that held up Wiles. Its proof follows from the work in Taylor-Wiles that establishes that it is enough to show that $\Re_Q \to \mathcal{T}_Q$ is an isomorphism for "sufficiently many" Q. This is then established by carefully picking the "nicest" sets Q, for which we can control somewhat the form of $\Re_Q$ and $\mathcal{T}_Q$, as follows.

Suppose that F is a finite field of odd characteristic ℓ and that $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(F)$ is a representation. We shall be interested in specifying various auxiliary primes q that will always be such that $q \equiv 1 \pmod{\ell}$, $\bar{\rho}$ is unramified at q, and $\bar{\rho}(Fr_q)$ has distinct eigenvalues all in F. Call such a prime *special.* We can control lifts ρ of $\bar{\rho}$ to rings A in $\mathcal{C}_F$ at special primes as follows.

Lemma 10.1 *With the above assumptions, there is a basis of*

$M := A^2$ on which $\rho(G_{\mathbf{Q}_q})$ acts diagonally. In fact,

$$\rho(I_q) = \begin{pmatrix} \chi_1 & 0 \\ 0 & \chi_2 \end{pmatrix}$$

with $\chi_1, \chi_2 : I_q \to A^*$ characters of order a power of ℓ that divides $q-1$.

Proof: By Hensel's lemma, there exists a basis of M on which $\rho(Fr_q)$ is diagonal, say $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. Let $x \in I_q$. Since $\bar{\rho}$ is unramified at q, $\rho(x) \in \Gamma_n(A)$. Since $\Gamma_n(A)$ is a pro-ℓ group whereas the 1st ramification subgroup G_1 of I_q is a pro-q group, $q \neq \ell$ forces $\rho(G_1) = \{1\}$. Thus ρ_q factors through $G_{\mathbf{Q}_q}/G_1$, whose structure we found earlier. In particular,

$$\rho(Fr_q)\rho(x)\rho(Fr_q)^{-1} = \rho(x)^q$$

Let $\rho(x) = 1 + (a_{ij})$ so that $a_{ij} \in \mathbf{m}$, the maximal ideal of A. Let I be the ideal of A generated by a_{ij} for $i \neq j$. The above equation implies that $\lambda_i a_{ij} \lambda_j^{-1} \equiv q a_{ij} \pmod{\mathbf{m}} I$. Our assumption that $\lambda_i \neq \lambda_j$ yields that $a_{ij} \in \mathbf{m}I$ for all $i \neq j$. Thus, $I = \mathbf{m}I$, so by Nakayama $I = (0)$. Hence, $a_{ij} = 0$ for $i \neq j$, i.e. $\rho(x)$ is diagonal.

By the equation above again, χ_i is of finite order dividing $q-1$. Since $\chi_i \equiv 1 \pmod{\mathbf{m}}$, these orders are powers of ℓ. QED

Letting Δ_q be the largest quotient of $(\mathbf{Z}/q)^*$ of order a power of ℓ, this is naturally a quotient of I_q via the cyclotomic character. The last result says that each χ_i factors through Δ_q.

We shall be considering sets Q of primes q satisfying the conditions above. Letting $\Delta_Q = \prod_{q \in Q}$, choosing one of the χ_i for each $q \in Q$ defines a homomorphism $\Delta_Q \to A^*$. In particular, if we take A to be the universal type Q deformation ring $\Re_Q$ or the universal modular type Q deformation ring $\mathcal{T}_Q$, then this makes them $W(F)[\Delta_Q]$-algebras. Letting I_Q be the augmentation ideal of $W(F)[\Delta_Q]$, our above lemma nicely relates $\Re_Q$ and $\Re_\emptyset$.

Theorem 10.2 *A lift of $\bar{\rho}$ of type Q to A in $\mathcal{C}_F$ is unramified at all $q \in Q$ if and only if the kernel of the representing map $\Re_Q \to A$ contains $I_Q \Re_Q$.*

Thus the natural map $\Re_Q \to \Re_\emptyset$ induces an isomorphism $\Re_Q / I_Q \Re_Q \to \Re_\emptyset$.

Proof: First, note that in the above lemma, $\chi_2 = \chi_1^{-1}$ since the determinant of ρ lands in both $\Gamma_1(A)$ and in $\mathbf{Z}_\ell^*$ and has finite order, so is trivial. Thus, one character -call it χ_q - determines the other. A lift is unramified at q if and only if this character is trivial. Noting that I_Q is generated by the $\delta - 1 (\delta \in \Delta_Q)$, so by the $\delta - 1 (\delta \in \Delta_q, q \in Q)$, we just need to show that $\chi_q = 1$ if and only if $(\delta - 1)x = 0 (x \in \Re_Q)$, but this latter is $\chi_q(\delta)x - x$, which $= 0$ if and only if $\chi_q(\delta) = 1$. QED

We shall write O for $W(F)$ (in fact, sometimes we shall want to increase $W(F)$ and this notation includes that possibility). If Q consists of r primes, consider the ring homomorphism

$$O[[S_1, ..., S_r]] \to O[\Delta_Q]$$

given by $1 + S_i \mapsto$ a chosen generator of the cyclic group Δ_{q_i}. This then induces an isomorphism

$$O[[S_1, ..., S_r]]/((1 + S_1)^{|\Delta_{q_1}|} - 1, ..., (1 + S_r)^{|\Delta_{q_r}|} - 1) \to O[\Delta_Q],$$

under which I_Q corresponds to $< S_1, ..., S_r >$. Via this map, we consider $\Re_Q$ and $\mathcal{T}_Q$ as $O[[S_1, ..., S_r]]$-algebras.

Theorem 10.3 *If the r primes in Q are special and also satisfy that*

$$H^1_\emptyset(G_{\mathbf{Q}}, \mathrm{Ad}^0(\bar\rho)^*) \xrightarrow{\prod res_q} \oplus_{r \in Q} H^1(G_{\mathbf{Q}_q}, \mathrm{Ad}^0(\bar\rho)^*)(\dagger)$$

is an isomorphism, then $\Re_Q$ is generated by r elements.

Proof: $\Re_Q$ is a quotient of $O[[T_1, ..., T_d]]$ where d is the dimension of $H^1_Q(G_{\mathbf{Q}}, \mathrm{Ad}^0(\bar\rho))$. So we need an upper bound (of r) on this dimension.

By Wiles' formula, looking at the qth factor, $q \in Q$, gives:

$$\frac{|L_q|}{|H^0(G_{\mathbf{Q}_q}, \mathrm{Ad}^0(\bar\rho))|} = \frac{|H^1(G_{\mathbf{Q}_q}, \mathrm{Ad}^0(\bar\rho))|}{H^0(G_{\mathbf{Q}_q}, \mathrm{Ad}^0(\bar\rho))|}$$

which $= |H^2(G_{\mathbf{Q}_q}, \mathrm{Ad}^0(\bar\rho))|$ since the Euler characteristic is 1, so by Tate duality $= |H^0(G_{\mathbf{Q}_q}, \mathrm{Ad}^0(\bar\rho)^*)|$, i.e. the order of the $G_{\mathbf{Q}_q}$-invariants of $ad^0(\bar\rho)^*$. Since $\bar\rho$ is unramified at q, this is the Fr_q-invariants, and an explicit computation gives that this is $|F|$.

We get such a term for each $q \in Q$ and so overall:

$$dim_F H^1_Q(G_{\mathbf{Q}}, \mathrm{Ad}^0(\bar\rho)) = r + dim_F H^1_Q(G_{\mathbf{Q}}, \mathrm{Ad}^0(\bar\rho)^*).$$

Since $\Re_Q/I_Q\Re_Q \cong \Re_\emptyset$, $\Re_Q$ is generated by the same number of elements as $\Re_\emptyset$, which the last argument gives as $dim H^1_\emptyset(G_{\mathbf{Q}}, \mathrm{Ad}^0(\bar\rho)^*)$. We just need to show this is r, but this follows from our hypothesis and the fact that $dim H^1(G_{\mathbf{Q}_q}, \mathrm{Ad}^0(\bar\rho)^*) = 1$. QED

Moving towards the hypotheses of theorem 8.13, we want for each $n \in \mathbf{Z}$ and $\psi \in H^1_Q(G_{\mathbf{Q}}, \mathrm{Ad}^0(\bar\rho)^*)$ a prime q depending on ψ such that

(1) $q \equiv 1 \pmod{\ell)^n}$

(2) q special

(3) $res_q(\psi) \neq 0$ (to ensure † holds).

Recall Chebotarev's theorem. This states that if $K/\mathbf{Q}$ is a finite Galois extension and $g \in Gal(K/\mathbf{Q})$, then there exist infinitely many primes q unramified in $K/\mathbf{Q}$ such that $Fr_q = g$ (at least up to conjugation). The above conditions then translate into finding $\sigma \in G_{\mathbf{Q}}$ satisfying

(1) $\sigma \in G_{\mathbf{Q}(\zeta_{\ell^n})} = ker(G_{\mathbf{Q}} \to Gal(\mathbf{Q}(\zeta_{\ell^n})/\mathbf{Q}))$ via the isomorphism $Gal(\mathbf{Q}(\zeta_{\ell^n})/\mathbf{Q})) \to (\mathbf{Z}/\ell^n)^*$

(2) $\mathrm{Ad}^0(\bar{\rho})(\sigma)$ has an eigenvalue $\neq 1$

(3) $\psi(\sigma) \notin (\sigma - 1)\mathrm{Ad}^0(\bar{\rho})^*$

where Ad^0 is the homomorphism $GL_2(F) \to Aut(M_2^0(F)) \cong GL_3(F)$ given by conjugation in $M_2(F)$.

Exercise: Show that if $\bar{\rho}(\sigma)$ has eigenvalues λ and μ, then $\mathrm{Ad}^0(\bar{\rho})(\sigma)$ has eigenvalues $1, \lambda/\mu, \mu/\lambda$. Deduce that the eigenvalues of $\bar{\rho}(\sigma)$ are distinct if and only if $\mathrm{Ad}^0(\bar{\rho})(\sigma)$ does not have 1 as an eigenvalue. The equivalence of both (2)'s above follows if we ensure F is large enough to contain all eigenvalues λ, μ.

Since the scalar matrices act trivially by conjugation, Ad^0 factors through $PGL_2(F)$ and so the image of $\mathrm{Ad}^0(\bar{\rho})$ restricted to $G_{\mathbf{Q}(\zeta_{\ell^n})}$ can be considered as a finite subgroup of $PGL_2(F)$ and so of $PGL_2(\bar{\mathbf{F}}_\ell)$. All such subgroups are explicitly known, namely up to conjugation a subgroup of the upper triangular matrices, $PSL_2(\mathbf{F}_{\ell^n}), PGL_2(\mathbf{F}_{\ell^n}), D_n, A_4, S_4, A_5$. The idea is to show that in each case

$$H^1(Gal(K/\mathbf{Q}), \mathrm{Ad}^0(\bar{\rho})^*) = 0$$

either by direct calculation(Cline-Parshall-Scott) or by using that $Gal(K/\mathbf{Q})$ has order prime to ℓ (Schur-Zassenhaus).

11

Putting it together - the final trick

Theorem 11.1 *Fermat's Last Theorem holds, i.e. if a, b, c, n are integers such that $a^n + b^n + c^n = 0$ and $n \geq 3$, then $abc = 0$.*

More generally, the argument below establishes:

Theorem 11.2 *Every semistable elliptic curve over $\mathbf{Q}$ is modular.*

Let E be the corresponding Frey elliptic curve. Consider its associated 3-division representation $\bar{\rho} : G_{\mathbf{Q}} \to GL_2(\mathbf{F}_3)$. There are two possibilities - either $\bar{\rho}$ is irreducible or reducible. Suppose first it is irreducible.

Lemma 11.3 *In this case, $\bar{\rho}$ restricted to $G_{\mathbf{Q}(\sqrt{-3})}$ is absolutely irreducible.*

Proof: Let H be the image of $\bar{\rho}$ in $PGL_2(\mathbf{F}_3) \cong S_4$. Suppose the lemma is false. Then $H \neq S_4$ else $\bar{\rho}$ would be surjective by the exercise that follows, and then the image of $G_{\mathbf{Q}(\sqrt{-3})}$ would be $SL_2(\mathbf{F}_3)$, which is absolutely irreducible.

We consider the subgroups of S_4 in turn, first noting that $H \subseteq A_4 \cong PSL_2(\mathbf{F}_3)$ is impossible since then $\det(\bar{\rho})$ would be trivial. Also, H cannot be a subgroup of any S_3 since then the image of $\bar{\rho}$ is in a Borel subgroup and so $\bar{\rho}$ is reducible, contradicting our hypothesis.

This leaves the only possibilities that H is dihedral of order 8 or a subgroup of index 2 of it. Since E is semistable at every prime $p \neq 3$, $\bar{\rho}(I_p)$ has order 1 or 3. Since $3 \not| |H|$, $\bar{\rho}(I_p)$ has order 1, i.e. $\bar{\rho}$ can only be ramified at the prime 3. Thus the abelianization H/H' is the Galois group of an abelian extension of $\mathbf{Q}$ of degree 4 ramified only at 3. By Kronecker-Weber, this has to be a subextension of some $\mathbf{Q}(\zeta_{3^r})$, but this has degree over $\mathbf{Q}$ not divisible by 4. QED

The hypotheses of the general case are now satisfied; namely we have a representation $\bar{\rho} : G_{\mathbf{Q}} \rightarrow GL_2(\mathbf{F}_\ell)$ ($\ell = 3$ in fact) such that

(i) $\bar{\rho}$ restricted to $G_K, K = \mathbf{Q}(\sqrt{(-1)^{(\ell-1)/2}\ell})$ is absolutely irreducible;

(ii) $\bar{\rho}$ is semistable (follows from theory of Tate curves);

(iii) $\bar{\rho}$ is modular (follows from Langlands-Tunnell).

Let Σ denote the set of primes that divide abc. Then $\rho_{3^\infty} : G_{\mathbf{Q}} \rightarrow GL_2(\mathbf{Z}_3)$ is a lift of $\bar{\rho}$ of type Σ. We showed in chapter 11 that $\Re_\Sigma \rightarrow \mathcal{T}_\Sigma$ is an isomorphism. This means that the universal map $\Re_\Sigma \rightarrow \mathbf{Z}_3$ producing ρ_{3^∞} actually factors through $\mathcal{T}_\Sigma$, whence ρ_{3^∞} is modular, i.e. there exists a cuspidal eigenform f such that $tr\rho_{3^\infty}(Fr_p) = a_p(f)$ for all but finitely many primes p.

Next consider $\rho_{n^\infty} : G_{\mathbf{Q}} \rightarrow GL_2(\mathbf{Z}_n)$, where n is the Fermat exponent. Since $tr\rho_{n^\infty}(Fr_p) = tr\rho_{3^\infty}(Fr_p) = a_p(f)$ for all but finitely many primes p, it follows that ρ_{n^∞} is modular (note Faltings' theorem is not needed) and so ρ_n is modular too. We showed, however, that ρ_n is good at all primes $\neq 2$ and semistable at 2, whence $N(\rho_n) = 2$. By Ribet, ρ_n is then associated to a cuspidal eigenform of level 2 and weight 2 and trivial Nebentypus, but there are none such. This contradiction concludes the proof except for the matter of dealing with the possibility of $\bar{\rho}$ being reducible.

This possibility also troubled Wiles. Unable to handle it, he had the rest of his work typed up in the spring of 1993 (with the amusing typo "Fermat's lost theorem"), when he spotted how an idea of Mazur would fix the problem. This is called the "3-5 switch". If Wiles had not spotted this, then Elkies' alternative approach [?] of twisting the curves would have finished off FLT.

Theorem 11.4 *Suppose* $\rho_3 : G_{\mathbf{Q}} \to GL_2(\mathbf{F}_3)$ *is reducible. Then* $\rho_5 : G_{\mathbf{Q}} \to GL_2(\mathbf{F}_5)$ *is irreducible.*

Proof: Suppose both are reducible. Then $E(\bar{\mathbf{Q}})$ has a $G_{\mathbf{Q}}$-stable subgroup of order 15. Under the 1-1 correspondence between $Y_0(N)(K)$ and isomorphism classes of elliptic curves over a field K together with a subgroup of order N defined over K, this elliptic curve corresponds to a point of $Y_0(15)(\mathbf{Q})$. Now, $X_0(15)$ has genus 1 and so is an elliptic curve (in fact $15A$ in Cremona's tables). Its rational points form a group of order 8. Of these 4 are cusps so not in $Y_0(15)(\mathbf{Q})$, whereas the other 4 correspond to elliptic curves of conductor 50, which are therefore not semistable. QED

A similar calculation to the first lemma above shows that ρ_5 restricted to $G_{\mathbf{Q}(\sqrt{5})}$ is absolutely irreducible. ρ_5 is also semistable since E is. The one thing missing is that we do not know that it is modular. Since $GL_2(\mathbf{F}_5)$ is nonsolvable, the Langlands-Tunnell approach fails and we take an alternative route.

Theorem 11.5 *("3-5 switch") There exists a semistable elliptic curve A over $\mathbf{Q}$ such that $A[5] \cong E[5]$ as $G_{\mathbf{Q}}$-modules and $A[3]$ is irreducible as a $G_{\mathbf{Q}}$-module.*

Once this is proven, then by the theorem proven so far, A is modular, so its ρ_5 is modular. But this is the same ρ_5 as for E. Thus the missing piece is filled in.

Proof: The idea is to consider the collection of all elliptic curves A such that $A[5] \cong E[5]$ as $G_{\mathbf{Q}}$-modules by an isomorphism π that makes the diagram of Weil pairings commute:

$$\begin{tikzcd}
A[5] \times A[5] \arrow[rr, "\pi\times\pi"] \arrow[dr] & & E[5] \times E[5] \arrow[dl] \\
& \mu_5 &
\end{tikzcd}$$

and note that this is in 1-1 correspondence with a curve Y' which is a twist of $Y(5)$, so of genus 0. Then we use Hilbert's irreducibility theorem. QED

References

[1] N. Boston. A Taylor-made plug for Wiles' proof. *College Mathematics Journal*, 26:100–105, 1995.

[2] N. Bourbaki. *Commutative algebra. Chapters 1–7.* Springer-Verlag, Berlin, 1998.

[3] J. Buhler, R. Crandall, R. Ernvall, and T. Metsankyla. Irregular primes and cyclotomic invariants to four million. *Math. Comp.*, 61, no. 203:151–153, 1993.

[4] J. Cassels and A. Fröhlich. *Algebraic number theory. Proceedings of the instructional conference held at the University of Sussex, Brighton, September 1–17, 1965.* Harcourt Brace Jovanovich, 1986.

[5] G. Cornell, J. Silverman, and G. Stevens. *Modular forms and Fermat's last theorem.* Springer-Verlag, 1997.

[6] H. Darmon, F. Diamond, and R. Taylor. Fermat's last

theorem. Elliptic curves, modular forms and Fermat's last theorem (Hong Kong, 1993):2–140, 1997.

[7] P. Deligne. Formes modulaires et représentations ℓ-adiques. *Séminaire Bourbaki*, pages 139–172, 1971.

[8] P. Deligne and J.-P. Serre. Formes modulaires de poids 1. *Annales de l'École Normale Superieure*, 7:507–530, 1974.

[9] F. Destrempes. Deformation of Galois representations: the flat case. *Seminar on Fermat's Last Theorem*, 17:209–231, 1995.

[10] F. Diamond. The refined conjecture of Serre. pages 22–37, 1995.

[11] F. Diamond and J. Im. Modular forms and modular curves. *Seminar on Fermat's Last Theorem*, 17:39–133, 1995.

[12] J.-M. Fontaine and B. Mazur. Geometric Galois representations. pages 41–78, 1995.

[13] F. Gouvea. A marvelous proof. *American Mathematical Monthly*, 101:203–222, 1994.

[14] F. Gouvea. Deformations of Galois representations. *Seminar on Fermat's Last Theorem*, 17, 1995.

[15] A. Granville and M. Monagan. The first case of fermat's last theorem is true for all prime exponents up to $714,591,416,091,389$. *Trans. AMS*, 306:329–359, 1988.

[16] K. Ireland and M. Rosen. *A classical introduction to modern number theory. Second edition.* Graduate Texts in Mathematics 84. Springer-Verlag, New York, 1990.

[17] N. Jacobson. *Basic Algebra,II.* W.H.Freeman, New York, 1989.

[18] U. Jannsen and K. Wingberg. Die struktur der absoluten galoisgruppe $\wp$-adischer zahlkörper. *Inv. Math.*, 70:71–98, 1982.

[19] A. Knapp. *Elliptic curves.* Princeton University Press, 1992.

[20] B. Mazur. Deforming Galois representations. Galois groups over $\mathbf{Q}$, series=MSRI Publications, 16, publisher=Springer-Verlag, pages=385–437, year=1989,.

[21] J.S. Milne. *Arithmetic duality theorems.* Perspectives in Mathematics 1. 1986.

[22] J. Oesterlé and J.-P. Serre. Travaux de Wiles (et Taylor, ...). *Astérisque*, 237:319–355, 1996.

[23] P. Ribenboim. *Fermat's last theorem for amateurs.* Springer-Verlag, New York, 1999.

[24] K. Ribet. On modular representations of $Gal(\overline{\mathbf{q}}/\mathbf{q})$ arising from modular forms. *Inv.Math.*, 100:431–476, 1990.

[25] K. Ribet. Galois representations and modular forms. *Bulletin of AMS*, 32:375–402, 1995.

[26] K. Rubin and A. Silverberg. A report on Wiles' Cambridge lectures. *Bulletin of AMS*, 31:15–38, 1994.

[27] M. Schlessinger. Functors of Artin rings. *Trans. AMS*, 130:208–222, 1968.

[28] J.-P. Serre. *A course in arithmetic.* Graduate Texts in Mathematics 7. Springer-Verlag, New York-Heidelberg, 1973.

[29] J.-P. Serre. *Local fields.* Graduate Texts in Mathematics 67. Springer-Verlag, New York-Berlin, 1979.

[30] J.-P. Serre. Sur les représentations modulaires de degré 2 de $Gal(\overline{\mathbf{q}}/\mathbf{q})$. *Duke Math. J.*, 454:179–230, 1987.

[31] G. Shimura. *Introduction to the arithmetic theory of automorphic functions.* Princeton University Press, 1971.

[32] J. Silverman. *Arithmetic of elliptic curves.* Graduate Texts in Mathematics 106. Springer-Verlag, New York, 1986.

[33] I. Stewart and D. Tall. *Algebraic number theory and Fermat's last theorem. Third edition.* A K Peters, Ltd., Natick, MA, 2002.

[34] R. Taylor and A. Wiles. Ring-theoretic properties of certain Hecke algebras. *Ann. of Math.*, 141:553–572, 1995.

[35] L. Washington. Galois cohomology. Modular forms and Fermat's last theorem:101–120, 1997.

[36] L. Washington. *Introduction to cyclotomic fields. Second edition.* Graduate Texts in Mathematics 83. Springer-Verlag, New York, 1997.

[37] W. Waterhouse. *Introduction to affine group schemes.* Graduate Texts in Mathematics 66. Springer-Verlag, New York, 1979.

[38] A. Wiles. Modular elliptic curves and Fermat's last theorem. *Ann. of Math.*, 141:443–551, 1995.