Fermat's last theorem

Henri Darmon
(DARMON@MATH.MCGILL.CA)
DEPARTMENT OF MATHEMATICS
MCGILL UNIVERSITY
MONTREAL, QC
CANADA H3A 2K6

Fred Diamond
(FDIAMOND@PMMS.CAM.AC.UK)
D.P.M.M.S.
CAMBRIDGE UNIVERSITY
CAMBRIDGE, CB2 1SB
UNITED KINGDOM

Richard Taylor
(TAYLORR@MATHS.OX.AC.UK)
MATHEMATICS INSTITUTE
OXFORD UNIVERSITY
24-29 ST. GILES
OXFORD, OX1 3LB
UNITED KINGDOM

The authors would like to give special thanks to N. Boston, K. Buzzard, and B. Conrad for providing so much valuable feedback on earlier versions of this paper. They are also grateful to A. Agboola, M. Bertolini, B. Edixhoven, J. Fearnley, B.H. Gross, R. Gross, L. Guo, F. Jarvis, H. Kisilevsky, E. Liverance, J. Manoharmayum, K. Ribet, D. Rohrlich, M. Rosen, R. Schoof, J.-P. Serre, C. Skinner, D. Thakur, J. Tilouine, J. Tunnell, A. Van der Poorten, and L.

Washington for their helpful comments.

Darmon thanks the members of CICMA and of the Quebec-Vermont Number Theory Seminar for many stimulating conversations on the topics of this paper, particularly in the Spring of 1995. For the same reason Diamond is grateful to the participants in an informal seminar at Columbia University in 1993-94, and Taylor thanks those attending the Oxford Number Theory Seminar in the Fall of 1995.

Parts of this paper were written while the authors held positions at other institutions: Darmon at Princeton University, Diamond at the Institute for Advanced Study, and Taylor at Cambridge University. During some of the period, Diamond enjoyed the hospitality of Princeton University, and Taylor that of Harvard University and MIT. The writing of this paper was also supported by research grants from NSERC (Darmon), EPSRC #GR/J94761 and NSF # DMS 9304580 (Diamond) and by an advanced fellowship from EPSRC (Taylor).

This article owes everything to the ideas of Wiles, and the arguments presented here are fundamentally his [W3], though they include both the work [TW] and several simplifications to the original arguments, most notably that of Faltings. In the hope of increasing clarity, we have not always stated theorems in the greatest known generality, concentrating instead on what is needed for the proof of the Shimura-Taniyama conjecture for semi-stable elliptic curves. This article can serve as an introduction to the fundamental papers [W3] and [TW], which the reader is encouraged to consult for a different, and often more in-depth, perspective on the topics considered. Another useful more advanced reference is the article [Di2] which strengthens the methods of [W3] and [TW] to prove that every elliptic curve that is semistable at 3 and 5 is modular.

Introduction

Fermat's Last Theorem

Fermat's Last Theorem states that the equation

$$x^n + y^n = z^n, \quad xyz \neq 0$$

has no integer solutions when n is greater than or equal to 3. Around 1630, Pierre de Fermat claimed that he had found a "truly wonderful" proof of this theorem, but that the margin of his copy of Diophantus' *Arithmetica* was too small to contain it:

> "Cubum autem in duos cubos, aut quadrato quadratum in duos quadrato quadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duos ejusdem nominis fas est dividere; cujus rei demonstrationem mirabile sane detexi. Hanc marginis exiguitas non caperet."

Among the many challenges that Fermat left for posterity, this was to prove the most vexing. A tantalizingly simple problem about whole numbers, it stood unsolved for more than 350 years, until in 1994 Andrew Wiles finally laid it to rest.

Prehistory: The only case of Fermat's Last Theorem for which Fermat actually wrote down a proof is for the case $n = 4$. To do this, Fermat introduced the idea of *infinite descent* which is still one the main tools in the study of Diophantine equations, and was to play a central role in the proof of Fermat's Last Theorem 350 years later. To prove his Last Theorem for exponent 4, Fermat showed something slightly stronger, namely that the equation $x^4 + y^4 = z^2$ has no solutions in relatively prime integers with $xyz \neq 0$. Solutions to such an equation correspond to rational points on the elliptic curve $v^2 = u^3 - 4u$. Since every integer $n \geq 3$ is divisible either by an odd prime or by 4, the result of Fermat allowed one to reduce the study of Fermat's equation to the case where $n = \ell$ is an *odd prime*.

In 1753, Leonhard Euler wrote down a proof of Fermat's Last Theorem for the exponent $\ell = 3$, by performing what in modern language we would call a 3-descent on the curve $x^3 + y^3 = 1$ which is also an elliptic curve. Euler's argument (which seems to have contained a gap) is explained in [Edw], ch. 2, and [Dic1], p. 545.

It took mathematicians almost 100 years after Euler's achievement to handle the case $\ell = 5$; this was settled, more or less simultaneously, by Gustav Peter Lejeune Dirichlet [Dir] and Adrien Marie Legendre [Leg] in 1825. Their elementary arguments are quite involved. (Cf. [Edw], sec. 3.3.)

In 1839, Fermat's equation for exponent 7 also yielded to elementary methods, through the heroic efforts of Gabriel Lamé. Lamé's proof was even more intricate than the proof for exponent 5, and suggested that to go further, new theoretical insights would be needed.

The work of Sophie Germain: Around 1820, in a letter to Gauss, Sophie Germain proved that if ℓ is a prime and $q = 2\ell + 1$ is also prime, then Fermat's equation $x^\ell + y^\ell = z^\ell$ with exponent ℓ has no solutions (x, y, z) with $xyz \not\equiv 0 \pmod{\ell}$. Germain's theorem was the first really general proposition on Fermat's Last Theorem, unlike the previous results which considered the Fermat equation one exponent at a time.

The case where the solution (x, y, z) to $x^\ell + y^\ell = z^\ell$ satisfies $xyz \not\equiv 0 \pmod{\ell}$ was called the *first case* of Fermat's Last Theorem, and the case where ℓ divides xyz, the *second case*. It was realized at that time that the first case was generally easier to handle: Germain's theorem was extended, using similar ideas, to cases where $k\ell + 1$ is prime and k is small, and this led to a proof that there were no first case solutions to Fermat's equation with prime exponents $\ell \leq 100$, which in 1830 represented a significant advance. The division between first and second case remained fundamental in much of the later work on the subject. In 1977, Terjanian [Te] proved that if the equation $x^{2\ell} + y^{2\ell} = z^{2\ell}$ has a solution (x, y, z), then 2ℓ divides either x or y, i.e., "the first case of Fermat's Last Theorem is true for even exponents". His simple and elegant proof used only techniques that were available to Germain and her contemporaries.

The work of Kummer: The work of Ernst Eduard Kummer marked the beginning of a new era in the study of Fermat's Last Theorem. For the first time, sophisticated concepts of algebraic number theory and the theory of L-functions were brought to bear on a question that had until then been addressed only with elementary methods. While he fell short of providing a complete solution, Kummer made substantial progress. He showed how Fermat's Last Theorem is intimately tied to deep questions on class numbers of cyclotomic fields which are still an active subject of research. Kummer's approach relied on the factorization

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

of Fermat's equation over the ring $\mathbb{Z}[\zeta_\ell]$ generated by the ℓth roots of unity. One observes that the greatest common divisor of any two factors in the product on the left divides the element $(1 - \zeta_\ell)$, which is an element of norm ℓ. Since the product of these numbers is a perfect ℓ-th power, one is tempted to conclude that $(x + y), \ldots, (x + \zeta_\ell^{\ell-1} y)$ are each ℓ-th powers in the ring $\mathbb{Z}[\zeta_\ell]$ up to units in this ring, and up to powers of $(1 - \zeta_\ell)$. Such an inference would be valid if one were to replace $\mathbb{Z}[\zeta_\ell]$ by $\mathbb{Z}$, and is a direct consequence of *unique factorization* of integers into products of primes. We say that a ring R has property UF if every non-zero element of R is uniquely a product of primes, up to units. Mathematicians such as Lamé made attempts at proving Fermat's Last Theorem based on the mistaken assumption that the rings $\mathbb{Z}[\zeta_\ell]$ had property UF. Legend even has it that Kummer fell into this trap, although this story now has been discredited; see for example [Edw], sec. 4.1. In fact, property UF is far from being satisfied in general: one now

knows that the rings $\mathbb{Z}[\zeta_\ell]$ have property UF only for $\ell < 23$ (cf. [Wa], ch. 1).

It turns out that the full force of property UF is not really needed in the applications to Fermat's Last Theorem. Say that a ring R has property UF_ℓ if the following inference is valid:

$$ab = z^\ell, \text{ and } \gcd(a,b) = 1 \Rightarrow a \text{ and } b \text{ are } \ell\text{th powers up to units of } R.$$

If a ring R has property UF, then it also has property UF_ℓ, but the converse need not be true. Kummer showed that Fermat's last theorem was true for exponent ℓ if $\mathbb{Z}[\zeta_\ell]$ satisfied the property UF_ℓ (cf. [Wa]). The proof is far from trivial, because of difficulties arising from the units in $\mathbb{Z}[\zeta_\ell]$ as well as from the possible failure of property UF. (A number of Kummer's contemporaries, such as Cauchy and Lamé, seem to have overlooked both of these difficulties in their attempts to prove Fermat's Last Theorem.)

Kummer then launched a systematic study of the property UF_ℓ for the rings $\mathbb{Z}[\zeta_\ell]$. He showed that even if $\mathbb{Z}[\zeta_\ell]$ failed to have unique factorization, it still possessed unique factorization into prime *ideals*. He defined the *ideal class group* as the quotient of the group of fractional ideals by its subgroup consisting of principal ideals, and was able to establish the finiteness of this class group. The order of the class group of $\mathbb{Z}[\zeta_\ell]$, denoted h_ℓ, could be taken as a measure of the failure of the ring $\mathbb{Z}[\zeta_\ell]$ to satisfy UF. It was rather straightforward to show that if ℓ did not divide h_ℓ, then $\mathbb{Z}[\zeta_\ell]$ satisfied the property UF_ℓ. In this case, one called ℓ a *regular prime*. Kummer thus showed that Fermat's last theorem is true for exponent ℓ if ℓ is a regular prime.

He did not stop here. For it remained to give an efficient means of computing h_ℓ, or at least an efficient way of checking when ℓ divides h_ℓ. The class number h_ℓ can be factorized as a product

$$h_\ell = h_\ell^+ h_\ell^-,$$

where h_ℓ^+ is the class number of the real subfield $\mathbb{Q}(\zeta_\ell)^+$, and h_ℓ^- is defined as h_ℓ / h_ℓ^+. Essentially because of the units in $\mathbb{Q}(\zeta_\ell)^+$, the factor h_ℓ^+ is somewhat difficult to compute, while, because the units in $\mathbb{Q}(\zeta_\ell)^+$ generate the group of units in $\mathbb{Q}(\zeta_\ell)$ up to finite index, the term h_ℓ^- can be expressed in a simple closed form. Kummer showed that if ℓ divides h_ℓ^+, then ℓ divides h_ℓ^-. Hence, ℓ divides h_ℓ if and only if ℓ divides h_ℓ^-. This allowed one to avoid the difficulties inherent in the calculation of h_ℓ^+. Kummer then gave an elegant formula for h_ℓ^- by considering the Bernoulli numbers B_n, which are rational numbers defined by the formula

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

He produced an explicit formula for the class number h_ℓ^-, and concluded that if ℓ does not divide the numerator of B_{2i}, for $1 \leq i \leq (\ell - 3)/2$, then ℓ is regular, and conversely.

The conceptual explanation for Kummer's formula for h_ℓ^- lies in the work of Dirichlet on the analytic class number formula, where it is shown that h_ℓ^- can be expressed as a product of special values of certain (abelian) L-series

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

associated to odd Dirichlet characters. Such special values in turn can be expressed in terms of certain generalized Bernoulli numbers $B_{1,\chi}$, which are related to the Bernoulli numbers B_i via congruences mod ℓ. (For more details, see [Wa].)

These considerations led Kummer to initiate a deep study relating congruence properties of special values of L-functions and of class numbers, which was to emerge as a central concern of modern algebraic number theory, and was to reappear – in a surprisingly different guise – at the heart of Wiles' strategy for proving the Shimura-Taniyama conjecture.

Later developments: Kummer's work had multiple ramifications, and led to a very active line of enquiry pursued by many people. His formulae relating Bernoulli numbers to class numbers of cyclotomic fields were refined by Kenneth Ribet [R1], Barry Mazur and Andrew Wiles [MW], using new methods from the theory of modular curves which also play a central role in Wiles' more recent work. (Later Francisco Thaine [Th] reproved some of the results of Mazur and Wiles using techniques inspired directly from a reading of Kummer.) In a development more directly related to Fermat's Last Theorem, Wieferich proved that if ℓ^2 does not divide $2^{\ell-1} - 1$, then the first case of Fermat's Last Theorem is true for exponent ℓ. (Cf. [Ri], lecture VIII.)

There were many other refinements of similar criteria for Fermat's Last theorem to be true. Computer calculations based on these criteria led to a verification that Fermat's Last theorem is true for all odd prime exponents less than four million [BCEM], and that the first case is true for all $\ell \leq 8.858 \cdot 10^{20}$ [Su].

The condition that ℓ is a regular prime seems to hold heuristically for about 61% of the primes. (See the discussion on p. 63, and also p. 108, of [Wa], for example.) In spite of the convincing numerical evidence, it is still not known if there are infinitely many regular primes. Ironically, it is not too difficult to show that there are infinitely many irregular primes. (Cf. [Wa].)

Thus the methods introduced by Kummer, after leading to very strong results in the direction of Fermat's Last theorem, seemed to become mired in difficulties, and ultimately fell short of solving Fermat's conundrum[1].

Faltings' proof of the Mordell conjecture: In 1985, Gerd Faltings [Fa] proved the very general statement (which had previously been conjectured

[1] However, W. McCallum has recently introduced a technique, based on the method of Chabauty and Coleman, which suggests new directions for approaching Fermat's Last Theorem via the cyclotomic theory. An application of McCallum's method to showing the *second* case of Fermat's Last Theorem for regular primes is explained in [Mc].

by Mordell) that any equation in two variables corresponding to a curve of genus strictly greater than one had (at most) finitely many rational solutions. In the context of Fermat's Last Theorem, this led to the proof that for each exponent $n \geq 3$, the Fermat equation $x^n + y^n = z^n$ has at most finitely many integer solutions (up to the obvious rescaling). Andrew Granville [Gra] and Roger Heath-Brown [HB] remarked that Faltings' result implies Fermat's Last Theorem for a set of exponents of density one.

However, Fermat's Last Theorem was still not known to be true for an infinite set of prime exponents. In fact, the theorem of Faltings seemed ill-equipped for dealing with the finer questions raised by Fermat in his margin, namely of finding a complete list of rational points on *all* of the Fermat curves $x^n + y^n = 1$ simultaneously, and showing that there are no solutions on these curves when $n \geq 3$ except the obvious ones.

Mazur's work on Diophantine properties of modular curves: Although it was not realized at the time, the chain of ideas that was to lead to a proof of Fermat's Last theorem had already been set in motion by Barry Mazur in the mid seventies. The modular curves $X_0(\ell)$ and $X_1(\ell)$ introduced in section 1.2 and 1.5 give rise to another naturally occurring infinite family of Diophantine equations. These equations have certain systematic rational solutions corresponding to the cusps that are defined over $\mathbb{Q}$, and are analogous to the so-called "trivial solutions" of Fermat's equation. Replacing Fermat curves by modular curves, one could ask for a complete list of all the rational points on the curves $X_0(\ell)$ and $X_1(\ell)$. This problem is perhaps even more compelling than Fermat's Last Theorem: rational points on modular curves correspond to objects with natural geometric and arithmetic interest, namely, elliptic curves with cyclic subgroups or points of order ℓ. In [Maz1] and [Maz2], B. Mazur gave essentially a complete answer to the analogue of Fermat's Last Theorem for modular curves. More precisely, he showed that if $\ell \neq 2, 3, 5$ and 7, (i.e., $X_1(\ell)$ has genus > 0) then the curve $X_1(\ell)$ has no rational points other than the "trivial" ones, namely cusps. He proved analogous results for the curves $X_0(\ell)$ in [Maz2], which implied, in particular, that an elliptic curve over $\mathbb{Q}$ with square-free conductor has no rational cyclic subgroup of order ℓ over $\mathbb{Q}$ if ℓ is a prime which is strictly greater than 7. This result appeared a full ten years before Faltings' proof of the Mordell conjecture.

Frey's strategy: In 1986, Gerhard Frey had the insight that these constructions might provide a precise link between Fermat's Last Theorem and deep questions in the theory of elliptic curves, most notably the Shimura Taniyama conjecture. Given a solution $a^\ell + b^\ell = c^\ell$ to the Fermat equation of prime degree ℓ, we may assume without loss of generality that $a^\ell \equiv -1 \pmod 4$ and that $b^\ell \equiv 0 \pmod{32}$. Frey considered (following Hellegouarch, [He], p. 262; cf. also Kubert-Lang [KL], ch. 8, §2) the elliptic curve

$$E : y^2 = x(x - a^\ell)(x + b^\ell).$$

This curve is *semistable*, i.e., it has square-free conductor. Let $E[\ell]$ denote the group of points of order ℓ on E defined over some (fixed) algebraic closure $\bar{\mathbb{Q}}$ of $\mathbb{Q}$, and let L denote the smallest number field over which these points are defined. This extension appears as a natural generalization of the cyclotomic fields $\mathbb{Q}(\zeta_\ell)$ studied by Kummer. What singles out the field L for special attention is that it has *very little ramification*: using Tate's analytic description of E at the primes dividing abc, it could be shown that L was ramified only at 2 and ℓ, and that the ramification of L at these two primes was rather restricted. (See theorem 2.15 of section 2.2 for a precise statement.) Moreover, the results of Mazur on the curve $X_0(\ell)$ could be used to show that L is *large*, in the following precise sense. The space $E[\ell]$ is a vector space of dimension 2 over the finite field $\mathbb{F}_\ell$ with ℓ elements, and the absolute Galois group $G_{\mathbb{Q}} = \mathrm{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q})$ acts $\mathbb{F}_\ell$-linearly on $E[\ell]$. Choosing an $\mathbb{F}_\ell$-basis for $E[\ell]$, the action is described by a representation

$$\bar{\rho}_{E,\ell} : \mathrm{Gal}\,(L/\mathbb{Q}) \hookrightarrow GL_2(\mathbb{F}_\ell).$$

Mazur's results in [Maz1] and [Maz2] imply that $\bar{\rho}_{E,\ell}$ is *irreducible* if $\ell > 7$ (using the fact that E is *semi-stable*). In fact, combined with earlier results of Serre [Se6], Mazur's results imply that for $\ell > 7$, the representation $\bar{\rho}_{E,\ell}$ is surjective, so that $\mathrm{Gal}\,(L/\mathbb{Q})$ is actually isomorphic to $GL_2(\mathbb{F}_\ell)$ in this case.

Serre's conjectures: In [Se7], Jean-Pierre Serre made a careful study of mod ℓ Galois representations $\bar{\rho} : G_{\mathbb{Q}} \longrightarrow GL_2(\mathbb{F}_\ell)$ (and, more generally, of representations into $GL_2(k)$, where k is any finite field). He was able to make very precise conjectures (see section 3.2) relating these representations to modular forms mod ℓ. In the context of the representations $\bar{\rho}_{E,\ell}$ that occur in Frey's construction, Serre's conjecture predicted that they arose from modular forms (mod ℓ) of weight two and level two. Such modular forms, which correspond to differentials on the modular curve $X_0(2)$, do not exist because $X_0(2)$ has genus 0. Thus Serre's conjecture implied Fermat's Last Theorem. The link between fields with Galois groups contained in $GL_2(\mathbb{F}_\ell)$ and modular forms mod ℓ still appears to be very deep, and Serre's conjecture remains a tantalizing open problem.

Ribet's work: lowering the level: The conjecture of Shimura and Taniyama (cf. section 1.8) provides a direct link between elliptic curves and modular forms. It predicts that the representation $\bar{\rho}_{E,\ell}$ obtained from the ℓ-division points of the Frey curve arises from a modular form of weight 2, albeit a form whose level is quite large. (It is the product of all the primes dividing abc, where $a^\ell + b^\ell = c^\ell$ is the putative solution to Fermat's equation.) Ribet [R5] proved that, if this were the case, then $\bar{\rho}_{E,\ell}$ would *also* be associated with a modular form mod ℓ of weight 2 and level 2, in the way predicted by Serre's conjecture. This deep result allowed him to reduce Fermat's Last Theorem to the Shimura-Taniyama conjecture.

Wiles' work: proof of the Shimura-Taniyama conjecture: In [W3] Wiles proves the Shimura-Taniyama conjecture for semi-stable elliptic curves, providing the final missing step and proving Fermat's Last Theorem. After more than 350 years, the saga of Fermat's Last theorem has come to a spectacular end.

The relation between Wiles' work and Fermat's Last Theorem has been very well documented (see, for example, [R8], and the references contained therein). Hence this article will focus primarily on the breakthrough of Wiles [W3] and Taylor-Wiles [TW] which leads to the proof of the Shimura-Taniyama conjecture for semi-stable elliptic curves.

From elliptic curves to ℓ-adic representations: Wiles' opening gambit for proving the Shimura-Taniyama conjecture is to view it as part of the more general problem of relating two-dimensional Galois representations and modular forms. The Shimura-Taniyama conjecture states that if E is an elliptic curve over $\mathbb{Q}$, then E is modular. One of several equivalent definitions of modularity is that for some integer N there is an eigenform $f = \sum a_n q^n$ of weight two on $\Gamma_0(N)$ such that

$$\#E(\mathbb{F}_p) = p + 1 - a_p$$

for all but finitely primes p. (By an eigenform, here we mean a cusp form which is a normalized eigenform for the Hecke operators; see section 1 for definitions.)

This conjecture acquires a more Galois theoretic flavour when one considers the two dimensional ℓ-adic representation

$$\rho_{E,\ell} : G_{\mathbb{Q}} \longrightarrow GL_2(\mathbb{Z}_\ell)$$

obtained from the action of $G_{\mathbb{Q}}$ on the ℓ-adic Tate module of

$$E : \mathcal{T}_\ell E = \lim_{\leftarrow} E[l^n](\bar{\mathbb{Q}}).$$

An ℓ-adic representation ρ of $G_{\mathbb{Q}}$ is said to arise from an eigenform $f = \sum a_n q^n$ with integer coefficients a_n if

$$\mathrm{tr}\,(\rho(\mathrm{Frob}_p)) = a_p,$$

for all but finitely many primes p at which ρ is unramified. Here Frob_p is a Frobenius element at p (see section 2), and its image under ρ is a well-defined conjugacy class.

A direct computation shows that $\#E(\mathbb{F}_p) = p + 1 - \mathrm{tr}\,(\rho_{E,\ell}(\mathrm{Frob}_p))$ for all primes p at which $\rho_{E,\ell}$ is unramified, so that E is modular (in the sense defined above) if and only if for some ℓ, $\rho_{E,\ell}$ arises from an eigenform. In fact the Shimura-Taniyama conjecture can be generalized to a conjecture that every ℓ-adic representation, satisfying suitable local conditions, arises from a modular form. Such a conjecture was proposed by Fontaine and Mazur [FM].

Galois groups and modular forms

Viewed in this way, the Shimura-Taniyama conjecture becomes part of a much larger picture: the emerging, partly conjectural and partly proven correspondence between certain modular forms and two dimensional representations of $G_{\mathbb{Q}}$. This correspondence, which encompasses the Serre conjectures, the Fontaine-Mazur conjecture, and the Langlands program for GL_2, represents a first step toward a higher dimensional, non-abelian generalization of class field theory.

Two-dimensional representations of $G_{\mathbb{Q}}$: In the first part of this century, class field theory gave a complete description of $G_{\mathbb{Q}}^{\mathrm{ab}}$, the maximal (continuous) abelian quotient of $G_{\mathbb{Q}}$. In fact the Kronecker-Weber theorem asserts that $G_{\mathbb{Q}}^{\mathrm{ab}} \cong \prod_p \mathbb{Z}_p^{\times}$, and one obtains a complete description of all one-dimensional representations of $G_{\mathbb{Q}}$. In the second half of this century much attention has focused on attempts to understand the whole group $G_{\mathbb{Q}}$, or more precisely to describe all its representations. Although there has been a fair degree of success in using modular forms to construct representations of $G_{\mathbb{Q}}$, less is known about how exhaustive these constructions are. The major results in the latter direction along these lines are the work of Langlands [Ll2] and the recent work of Wiles ([W3] completed by [TW]). Both concern two-dimensional representations of $G_{\mathbb{Q}}$ and give significant evidence that these representations are parametrised (in a very precise sense) by certain modular forms. The purpose of this article is to describe both the proven and conjectural parts of this theory, give a fairly detailed exposition of Wiles' recent contribution and explain the application to Fermat's Last theorem. To make this description somewhat more precise let us distinguish three types of representation.

Artin representations and the Langlands-Tunnell theorem: Continuous representations $\rho : G_{\mathbb{Q}} \to GL_2(\mathbb{C})$ are called (two-dimensional) Artin representations. Such representations necessarily have finite image, and are therefore semi-simple. We restrict our attention to those which are irreducible. They are conjectured to be in bijection (in a precise way) with certain newforms (a special class of eigenforms). Those ρ which are odd (i.e. the determinant of complex conjugation is -1), should correspond to weight 1 holomorphic newforms. Those which are even should correspond to certain non-holomorphic (Maass) newforms. Two partial but deep results are known.

(a) (Deligne-Serre) If f is a holomorphic weight one newform then the corresponding Artin representation can be constructed ([DS]).

(b) (Langlands-Tunnell) If ρ is a two dimensional Artin representation with soluble image then the corresponding modular form exists ([Ll2] and [Tu]).

The proof of the latter result is analytic in nature, invoking the trace formula and the theory of L-functions.

ℓ-adic representations and the Fontaine-Mazur conjecture: By an ℓ-adic representation we shall mean any continuous representation $\rho : G_{\mathbb{Q}} \to GL_2(K)$ which is unramified outside a finite set of primes and where K is a finite extension of $\mathbb{Q}_\ell$ (generalizing slightly the notion of ℓ-adic representation that was introduced before). Given a holomorphic newform f one can attach to f a system of ℓ-adic representations, following Eichler, Shimura, Deligne and Serre. These ℓ-adic representations are called modular. The Fontaine-Mazur conjecture (see [FM]) predicts if ρ is an odd, irreducible, ℓ-adic representation whose restriction to the decomposition group at ℓ is well enough behaved, then ρ is modular. (The restriction on the behaviour of the representation on the decomposition group at ℓ is essential in this conjecture; it is not true that all odd, irreducible two dimensional ℓ-adic representation are modular.) Before Wiles' work almost nothing was known about this conjecture, except that certain very special cases could be deduced from the work of Hecke, Langlands and Tunnell.

Mod ℓ representations and Serre's conjecture: A mod ℓ representation is a continuous representation $\bar{\rho} : G_{\mathbb{Q}} \longrightarrow GL_2(\bar{\mathbb{F}}_\ell)$. For example if $E/\mathbb{Q}$ is an elliptic curve then the action of $G_{\mathbb{Q}}$ on the ℓ-division points of E gives rise to a mod ℓ representation $\bar{\rho}_{E,\ell}$ which is just the reduction modulo ℓ of $\rho_{E,\ell}$. One can use the work of Eichler, Shimura, Deligne and Serre to associate to each mod ℓ eigenform a mod ℓ representation of $G_{\mathbb{Q}}$. The mod ℓ representations which arise in this way are called modular. Serre has conjectured [Se7] that every odd (absolutely) irreducible mod ℓ representation is modular and should arise from a mod ℓ eigenform with certain very specific properties. This conjecture can be thought of as having two parts.

The first asserts that every odd irreducible mod ℓ representation is modular. About this very little is known. It is known for $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_2)$ by work of Hecke. It is also known for $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_3)$. This latter result is an application of the Langlands-Tunnell theorem using the two accidents that there is a section to the homomorphism $GL_2(\mathbb{Z}[\sqrt{-2}]) \twoheadrightarrow GL_2(\mathbb{F}_3)$ and that $GL_2(\mathbb{F}_3)$ is soluble. Partial results for $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_5)$ follow from Wiles' work.

Given a mod ℓ representation arising from a mod ℓ eigenform, the second part of Serre's conjecture predicts the minimal weight and level for that mod ℓ eigenform. Here the situation is much better. There has been a lot of work over the last decade (including ideas from Mazur, Ribet, Carayol and Gross) and the problem is nearly completely resolved (see [Di1]). As was pointed out earlier, Ribet's contribution [R5] implies that, if one can show that the Galois representation $\bar{\rho}_{E,\ell}$ arising from the (semi-stable) Frey curve attached to a solution of Fermat's equation with exponent ℓ is modular, then one can show that this representation does not exist—because it would be modular of weight two and level two— and hence one can deduce Fermat's

Last Theorem.

However we have seen that to show $\bar{\rho}_{E,\ell}$ is modular it suffices to show that for some ℓ_0, the ℓ_0-adic representation ρ_{E,ℓ_0} is modular. In particular it suffices to verify that either $\rho_{E,3}$ or $\rho_{E,5}$ is modular. Hence the Shimura-Taniyama conjecture can be reduced to (part of) the Fontaine-Mazur conjecture for $\ell = 3$ and 5. We have seen that for these primes part of Serre's conjecture is known, so it turns out it suffices to prove results of the form "Serre's conjecture for ℓ implies the Fontaine-Mazur conjecture for ℓ". This is the direction of Wiles' work, although nothing quite this general has been proven yet.

Deformation theory: Thus the problem Wiles faces is to show that if ρ is an odd ℓ-adic representation which has irreducible modular reduction $\bar{\rho}$ and which is sufficiently well behaved when restricted to the decomposition group at ℓ, then ρ is modular. In fact he only proves a weakened version of such a result, but one which is sufficient to conclude that all semistable elliptic curves are modular.

Wiles approaches the problem by putting it in a more general setting. On the one hand he considers lifts of $\bar{\rho}$ to representations over complete noetherian local $\mathbb{Z}_\ell$-algebras R. For each finite set of primes Σ, one can consider lifts of type Σ; these are lifts which are well-behaved on a decomposition group at ℓ, and whose ramification at primes not in Σ is rather restricted. In particular, such a lift is unramified outside $\Sigma \cup S$ where S is the set of ramified primes of $\bar{\rho}$. A method of Mazur (see [Maz3]) can then be used to show that if $\bar{\rho}$ is absolutely irreducible, then there is a representation

$$\rho_\Sigma^{\text{univ}} : G_{\mathbb{Q}} \longrightarrow GL_2(R_\Sigma)$$

which is universal in the following sense. If $\rho : G_{\mathbb{Q}} \to GL_2(R)$ is a lift of $\bar{\rho}$ of type Σ, then there is a unique local homomorphism $R_\Sigma \longrightarrow R$ such that ρ is equivalent to the pushforward of $\rho_\Sigma^{\text{univ}}$. Thus the equivalence classes of type Σ lifts to $GL_2(R)$ can be *identified* with $\text{Hom}(R_\Sigma, R)$. The local ring R_Σ is called the *universal deformation ring* for representations of type Σ.

On the other hand Wiles constructs a candidate for a universal modular lifting of type Σ

$$\rho_\Sigma^{\text{mod}} : G_{\mathbb{Q}} \longrightarrow GL_2(\mathbb{T}_\Sigma).$$

The ring $\mathbb{T}_\Sigma$ is constructed from the algebra of Hecke operators acting on a certain space of modular forms. The universal property of R_Σ gives a map $R_\Sigma \to \mathbb{T}_\Sigma$. The problem thus becomes: to show that this map is an isomorphism[2]. In fact, it can be shown to be a surjection without great difficulty, and the real challenge is to prove injectivity, i.e., to show, in essence, that R_Σ is not larger than $\mathbb{T}_\Sigma$.

[2]Maps of this kind were already considered in [Maz3] and [BM], and it is conjectured in [MT] that these maps are isomorphisms in certain cases, though not in exactly the situations considered by Wiles.

By an ingenious piece of commutative algebra, Wiles found a numerical criterion for this map to be an isomorphism, and for the ring $\mathbb{T}_\Sigma$ to be a local complete intersection. This numerical criterion seems to be very close to a special case of the Bloch-Kato conjecture [BK]. Wiles further showed (by combining arguments from Galois cohomology and from the theory of congruences between modular forms) that this numerical criterion was satisfied if the minimal version $\mathbb{T}_\emptyset$ of this Hecke algebra (obtained by taking $\Sigma = \emptyset$, i.e., allowing the least possible amount of ramification in the deformations) was a complete intersection. Finally in [TW] it was proved that $\mathbb{T}_\emptyset$ is a complete intersection.

Outline of the paper

Chapter 1 recalls some basic notions from the classical theory of elliptic curves and modular forms, such as modular forms and modular curves over $\mathbb{C}$ and $\mathbb{Q}$, Hecke operators and q-expansions, and Eichler-Shimura theory. The Shimura-Taniyama conjecture is stated precisely in section 1.8.

Chapter 2 introduces the basic theory of representations of $G_\mathbb{Q}$. We describe Mazur's deformation theory and begin our study of the universal deformation rings using techniques from Galois cohomology and from the theory of finite flat group schemes. We also recall some basic properties of elliptic curves, both to explain Frey's argument precisely and illustrate the uses of ℓ-adic representations.

Chapter 3 explains how to associate Galois representations to modular forms. We then describe what was known and conjectured about associating modular forms to Galois representations before Wiles' work. After introducing the universal modular lifts of certain mod ℓ representations, we give the proof of Wiles' main theorems, taking for granted certain results of a more technical nature that are proved in the last two chapters.

Chapter 4 explains how to prove the necessary results concerning the structure of Hecke algebras: the generalization by Taylor and Wiles of a result of de Shalit, and the generalization by Wiles of a result of Ribet.

Chapter 5 establishes the fundamental results from commutative algebra discovered by Wiles, following modifications of the approach of Wiles and Taylor-Wiles proposed by Faltings and Lenstra.

Contents

1 Elliptic curves and modular forms

1.1 Elliptic curves

We begin with a brief review of elliptic curves. A general reference for the results discussed in this section is [Si1] and [Si2].

An elliptic curve E over a field F is a smooth projective curve over F of genus one with a distinguished F-rational point. If E/F is an elliptic curve and if ω is a non-zero holomorphic differential on E/F then E can be realised in the projective plane by an equation (called a Weierstrass equation) of the form

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3 \qquad \text{(W)}$$

such that the distinguished point is $(0:1:0)$ (sometimes denoted ∞ because it corresponds to the "point at infinity" in the affine model obtained by setting $Z = 1$) and $\omega = \frac{dx}{2y+a_1x+a_3}$. We also define the following quantities associated to (W):

$$\begin{aligned}
&b_2 = a_1^2 + 4a_2 \qquad b_4 = 2a_4 + a_1a_3 \qquad b_6 = a_3^2 + 4a_6\\
&b_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2\\
&\Delta = 9b_2b_4b_6 - b_2^2b_8 - 8b_4^3 - 27b_6^2\\
&j = (b_2^2 - 24b_4)^3/\Delta.
\end{aligned}$$

One can check that the equation (W) defines an elliptic curve if and only if Δ is nonzero. One can also check that such equations define elliptic curves which are isomorphic over $\bar{F}$ if and only if they give the same quantity j. Thus j only depends on E so we will denote it j_E. The quantity Δ depends only on the pair (E,ω) so we shall denote it $\Delta(E,\omega)$. If u belongs to $F^\times$ then $u^{12}\Delta(E,u\omega) = \Delta(E,\omega)$.

An elliptic curve E/F has a natural structure of a commutative algebraic group with the distinguished F-rational point as the identity element.

An algebraic map between two elliptic curves which sends the distinguished point of one to the distinguished point of the other is automatically a morphism of algebraic groups. A map between elliptic curves which has finite kernel (and hence, is generically surjective) is called an *isogeny*.

Elliptic curves over $\mathbb{C}$: If $F = \mathbb{C}$, then the curve E is isomorphic as a complex analytic manifold to the complex torus $\mathbb{C}/\Lambda$, where Λ is a lattice in $\mathbb{C}$, i.e., a discrete $\mathbb{Z}$-submodule of $\mathbb{C}$ of rank 2. The group law on $E(\mathbb{C})$

corresponds to the usual addition in $\mathbb{C}/\Lambda$. In terms of Λ, an affine equation for E in $A^2(\mathbb{C})$ is given by

$$y^2 = 4x^3 + g_2 x + g_3,$$

where

$$g_2 = -60 \sum_{\lambda \in \Lambda - \{0\}} \frac{1}{z^4}, \qquad g_3 = -140 \sum_{\lambda \in \Lambda - \{0\}} \frac{1}{z^6}.$$

In terms of this equation, the map from $\mathbb{C}/\Lambda$ to $E(\mathbb{C})$ sends z to $(x, y) = (\wp(z), \wp'(z))$, where $\wp(z)$ is the *Weierstrass $\wp$-function* associated to the lattice Λ. (Cf. [Si1], ch. VI.) The inverse map is given by integrating the holomorphic differential ω, i.e., sending $P \in E(\mathbb{C})$ to the image of $\int_\gamma \omega$ in $\mathbb{C}/\Lambda$, where γ is any path on $E(\mathbb{C})$ from ∞ to P, and Λ is the lattice of periods $\int_\gamma \omega$, where γ ranges over the integral homology $H_1(E(\mathbb{C}), \mathbb{Z})$. Replacing ω by $u\omega$ changes Λ to $u\Lambda$, so that Λ is determined by E only up to homotheties. We scale Λ so that one of its $\mathbb{Z}$-generators is 1, and another, τ, has strictly positive imaginary part. This gives the analytic isomorphism:

$$E(\mathbb{C}) \simeq \mathbb{C}/\langle 1, \tau \rangle.$$

The complex number τ in the complex upper half plane $\mathcal{H}$ is well defined, modulo the natural action of $SL_2(\mathbb{Z})$ on $\mathcal{H}$ by Möbius transformations. (Thus the set of isomorphism classes of elliptic curves over $\mathbb{C}$ can be *identified* with the quotient $\mathcal{H}/SL_2(\mathbb{Z})$.)

The map $z \mapsto e^{2\pi i z}$ identifies $\mathbb{C}/\langle 1, \tau \rangle$ with $\mathbb{C}^\times/q^{\mathbb{Z}}$, where $q = e^{2\pi i \tau}$ is the *multiplicative Tate period.* The analytic isomorphism

$$E(\mathbb{C}) \simeq \mathbb{C}^\times/q^{\mathbb{Z}}$$

has the virtue of generalizing to the p-adic setting in certain cases, as we will see shortly.

Note that $|q| < 1$. The invariant j can be expressed in terms of q by a convergent power series with *integer* coefficients:

$$j = q^{-1} + 744 + 196884q + \cdots. \tag{1.1.1}$$

The following basic facts are a direct consequence of the analytic theory:

Proposition 1.1 *The subgroup $E[n](\mathbb{C})$ of points of order n on $E(\mathbb{C})$ is isomorphic (non-canonically) to $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. More generally, if F is any field of characteristic zero, the subgroup $E[n](F)$ is contained in $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$.*

Proof. The analytic theory shows that $E(\mathbb{C})$ is isomorphic as an abstract group to a product of two circle groups, and the first statement follows. The second statement follows from the Lefschetz principle (cf. [Si1], ch. VI, §6). □

Proposition 1.2 *The endomorphism ring* $\mathrm{End}_{\mathbb{C}}(E)$ *of an elliptic curve over* $\mathbb{C}$ *is isomorphic either to* $\mathbb{Z}$ *or to an order in a quadratic imaginary field. The same is true if one replaces* $\mathbb{C}$ *by any field of characteristic* 0.

Proof: An endomorphism of $E(\mathbb{C}) \simeq \mathbb{C}/\Lambda$ induces multiplication by complex number α on the tangent space. Hence $\mathrm{End}_{\mathbb{C}}(E)$ is isomorphic to the ring of $\alpha \in \mathbb{C}$ satisfying $\alpha\Lambda \subset \Lambda$. Such a ring is isomorphic either to $\mathbb{Z}$ or to a quadratic imaginary order. The corresponding statement for fields of characteristic 0 follows as in the proof of proposition 1.1. □

If $\mathrm{End}_{\mathbb{C}}(E) \otimes \mathbb{Q}$ is a quadratic imaginary field, we say that E has *complex multiplication.*

Remark 1.3 It follows from the arithmetic theory of complex multiplication (cf. [Si2], ch. 1) that any elliptic curve E with complex multiplication is defined over an abelian extension of the quadratic imaginary field $K = \mathrm{End}_{\mathbb{C}}(E) \otimes \mathbb{Q}$. If E is defined over $\mathbb{Q}$, then K has class number one. There are only finitely many elliptic curves over $\mathbb{Q}$ with complex multiplication, up to "twists" (i.e., $\mathbb{C}$-isomorphism).

Elliptic curves over $\mathbb{Q}_p$: Now suppose that E is an elliptic curve defined over the p-adic field $\mathbb{Q}_p$. There is an equation

$$(W^{\min}) \qquad Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$$

for E with the property $a_i \in \mathbb{Z}_p$ for all i and $|\Delta|$ is minimal amongst all such equations for E. Although $(W^{\min})$ is not unique, the associated discriminant depends only on E and is denoted $\Delta_E^{\min}$. Moreover the reduction of $(W^{\min})$ modulo the uniformizer p defines a projective curve $\bar{E}$, which is independent of the particular minimal equation chosen. If (W) is any equation for E with coefficients in $\mathbb{Z}_p$ and with discriminant Δ, then $\Delta_E^{\min}$ divides Δ.

If $\bar{E}$ is a smooth curve we say that E has *good reduction* at p. If $\bar{E}$ has a unique singular point which is a node we say that E has *multiplicative reduction* at p. Otherwise $\bar{E}$ has a unique singular point which is a cusp and we say that E has *additive reduction* at p. If E has good or multiplicative reduction we say that it has *semi-stable* reduction at p, or simply that E is *semi-stable.*

If (W) defines a smooth curve mod p then E has good reduction at p and (W) is a minimal equation. If $\Delta \equiv 0 \bmod p$ but $b_2^2 \neq 24b_4 \bmod p$, then modulo p the equation (W) defines a curve with a node. In this case E has multiplicative reduction at p and (W) is a minimal equation.

Curves with good reduction: In that case p does not divide $\Delta_E^{\min}$, and the reduction $\bar{E}$ is an elliptic curve over $\mathbb{F}_p$.

If q is any power of p, and $\mathbb{F}_q$ is the field with q elements, we define the integer N_q to be the number of solutions to the equation $(W^{\min})$ in the

projective plane $\mathbb{P}^2(\mathbb{F}_q)$. Thus N_q is the order of the finite group $\bar{E}(\mathbb{F}_q)$. We define the integer a_q by the formula

$$a_q = q + 1 - N_q.$$

The integers a_q are completely determined by a_p: more precisely, we have

$$(1 - a_p p^{-s} + p^{1-2s})^{-1} = 1 + a_p p^{-s} + a_{p^2} p^{-2s} + a_{p^3} p^{-3s} + \cdots . \tag{1.1.2}$$

We call the expression on the left the (local) L-function associated to E over $\mathbb{Q}_p$, and denote it by $L(E/\mathbb{Q}_p, s)$. Concerning the size of a_p we have the following fundamental result of Hasse, whose proof can be found in [Si1], ch. V, § 1:

Theorem 1.4 $|a_p| \leq 2\sqrt{p}$.

A further division among curves of good reduction plays a significant role in our later discussion. We say that E has (good) *ordinary* reduction if p does not divide a_p, and that it has *supersingular* reduction if p divides a_p.

When E has good reduction at p, we define its local conductor at p to be $m_p(E) = 0$.

Curves of multiplicative reduction: Elliptic curves over $\mathbb{Q}_p$ which have multiplicative reduction at p can be understood by using the p-adic analytic description discovered by Tate. More precisely, we can formally invert the power series (1.1.1) expressing j in terms of q, to obtain to a power series for q in j^{-1}, having *integer* coefficients:

$$q = j^{-1} + 744j^{-2} + 750420j^{-3} + 872769632j^{-4} + \cdots . \tag{1.1.3}$$

If E has multiplicative reduction, then $j \in \mathbb{Q}_p$ is non-integral, and hence the power series (1.1.3) converges, yielding a well-defined value of q in $p\mathbb{Z}_p$. This is called Tate's p-adic period associated to E over $\mathbb{Q}_p$. Note that we have $v_p(q) = -v_p(j) = v_p(\Delta_E^{min})$.

We say that E has *split* (resp. *non-split*) multiplicative reduction at p if the two tangent lines to the node on $\bar{E}(\mathbb{F}_p)$ have slopes defined over $\mathbb{F}_p$ (resp. $\mathbb{F}_{p^2}$).

Proposition 1.5 (Tate) *There is a p-adic analytic isomorphism*

$$\Phi : \bar{\mathbb{Q}}_p^{\times} / q^{\mathbb{Z}} \longrightarrow E(\bar{\mathbb{Q}}_p),$$

which has the property that

$$\sigma(\Phi(x)) = \Phi(\sigma x^{\delta(\sigma)}), \quad \forall \sigma \in G_{\mathbb{Q}_p},$$

where $\delta : G_{\mathbb{Q}_p} \longrightarrow \pm 1$ is

- *the trivial character, if E has split multiplicative reduction;*
- *the unique unramified quadratic character of $G_{\mathbb{Q}_p}$, if E has non-split multiplicative reduction.*

The proof of this proposition is explained in [Si2], ch. V, for example.

We define the L-function $L(E/\mathbb{Q}_p, s)$ to be

$$L(E/\mathbb{Q}_p, s) = \begin{cases} (1-p^{-s})^{-1} & \text{if } E \text{ has split reduction,} \\ (1+p^{-s})^{-1} & \text{if } E \text{ has non-split reduction.} \end{cases} \tag{1.1.4}$$

In both cases the conductor $m_p(E)$ is defined to be 1.

Curves of additive reduction: If E has additive reduction at p, we simply define

$$L(E/\mathbb{Q}_p, s) = 1. \tag{1.1.5}$$

The conductor $m_p(E)$ is defined to be 2, if $p > 3$. When $p = 2$ or 3, it is determined by a somewhat more complicated recipe, given in [Ta].

Elliptic curves over $\mathbb{Q}$: Let E be an elliptic curve defined over $\mathbb{Q}$. In particular E may be viewed as a curve over $\mathbb{Q}_p$ for every p, and we define its (global) conductor by

$$N_E = \prod_p p^{m_p(E)}.$$

The curve E is said to be *semi-stable* if it is semi-stable over all p-adic fields $\mathbb{Q}_p$. Note that E is semi-stable if and only if its conductor N_E is square-free.

Using the fact that $\mathbb{Q}$ has class number 1, one can show that E has a *global minimal Weierstrass model* ($W^{\min}$) which gives the equation of a minimal Weierstrass model over each $\mathbb{Q}_p$. The associated discriminant, denoted $\Delta_E^{\min}$, depends only on E. The associated differential, denoted ω_E^{Neron}, is called the *Néron differential*. It is well-defined up to sign.

The following, known as the Mordell-Weil theorem, is the fundamental result about the structure of the group of rational points $E(\mathbb{Q})$. (Cf. For example [Si1].)

Theorem 1.6 *The group $E(\mathbb{Q})$ is a finitely generated abelian group. Hence*

$$E(\mathbb{Q}) \simeq T \oplus \mathbb{Z}^r,$$

where T is the (finite) torsion subgroup of $E(\mathbb{Q})$, and $r \geq 0$ is the rank *of E over $\mathbb{Q}$.*

Concerning the possible structure of T, there is the following deep result of Mazur, a variant of which also plays a crucial role in the proof of Fermat's Last Theorem:

Theorem 1.7 *If $E/\mathbb{Q}$ is an elliptic curve, then its torsion subgroup is isomorphic to one of the following possibilities:*

$$\mathbb{Z}/n\mathbb{Z}, \quad 1 \leq n \leq 10, \quad n = 12, \qquad \mathbb{Z}/2n\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}, \quad 1 \leq n \leq 4.$$

The proof is given in [Maz1] (see also [Maz2]). Thanks to this result, the structure of the torsion subgroup T is well understood. (Recently, the techniques of Mazur have been extended by Kamienny [Kam] and Merel [Mer] to prove uniform boundedness results on the torsion of elliptic curves over general number fields.)

Much more mysterious is the behaviour of the rank r. It is not known if r can be arbitrarily large, although results of Mestre [Mes] and Nagao [Na] show that it is greater or equal to 13 for infinitely many elliptic curves over $\mathbb{Q}$. It turns out that many of the deep results on $E(\mathbb{Q})$ and on r are based on the relation with L-functions.

We define the global L-function of the complex variable s by:

$$L(E/\mathbb{Q}, s) = \prod_p L(E/\mathbb{Q}_p, s). \tag{1.1.6}$$

Exercise 1.8 Using theorem 1.4, show that the infinite product defining the L-function $L(E/\mathbb{Q}, s)$ converges absolutely on the right half plane $\text{Real}(s) > 3/2$.

Conjecture 1.9 (Birch-Swinnerton-Dyer) *The L-function $L(E/\mathbb{Q}, s)$ has an analytic continuation to the entire complex plane,and in particular is analytic at $s = 1$. Furthermore:*

$$\text{ord}_{s=1} L(E/\mathbb{Q}, s) = r.$$

There is also a more precise form of this conjecture, which expresses the leading coefficient of $L(E/\mathbb{Q}, s)$ at $s = 1$ in terms of certain arithmetic invariants of $E/\mathbb{Q}$. For more details, see [Si1], conj. 16.5.

As we will explain in more detail in section 1.8, the analytic continuation of $L(E/\mathbb{Q}, s)$ now follows from the work of Wiles and Taylor-Wiles and a strengthening by Diamond [Di2], for a very large class of elliptic curves over $\mathbb{Q}$, which includes all the semi-stable ones.

Abelian varieties: Elliptic curves admit higher-dimensional analogues, called *abelian varieties*, which also play a role in our discussion. Analytically, the set of complex points on an abelian variety is isomorphic to a quotient $\mathbb{C}^g/\Lambda$, where Λ is a lattice in $\mathbb{C}^g$ of rank $2g$, satisfying the so-called Riemann period relations. A good introduction to the basic theory of abelian varieties can be found in [CS] and [We1].

1.2 Modular curves and modular forms over $\mathbb{C}$

Modular curves: The group $SL_2(\mathbb{Z})$ of two by two integer matrices of determinant one acts by fractional linear (Möbius) transformations on the complex upper half plane

$$\mathcal{H} = \{z \in \mathbb{C} \mid \mathrm{Im}\,(z) > 0\},$$

equipped with its standard complex analytic structure. The *principal congruence group* $\Gamma(N)$ of level N is the subgroup of matrices in $SL_2(\mathbb{Z})$ which reduce to the identity matrix modulo the positive integer N. A subgroup Γ of $SL_2(\mathbb{Z})$ is called a *congruence group* if it contains $\Gamma(N)$ for some N. The level of Γ is the smallest N for which this is true. The most important examples of congruence groups are:

- The group $\Gamma_0(N)$ consisting of all matrices that reduce modulo N to an upper triangular matrix.
- The group $\Gamma_1(N)$ consisting of all matrices that reduce modulo N to a matrix of the form $\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$.
- The principal congruence group $\Gamma(N)$ of level N consisting of all matrices that reduce modulo N to the identity.

Notice the natural inclusions of normal subgroups $\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N)$. The quotient $\Gamma_0(N)/\Gamma_1(N)$ is canonically isomorphic to $(\mathbb{Z}/N\mathbb{Z})^\times$ via

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto d \bmod N.$$

For any subgroup H of $(\mathbb{Z}/N\mathbb{Z})^\times$, we denote by $\Gamma_H(N)$ the group of matrices in $\Gamma_0(N)$ whose image in $\Gamma_0(N)/\Gamma_1(N)$ belongs to H.

If Γ is a congruence subgroup of $SL_2(\mathbb{Z})$, define Y_Γ to be the quotient of the upper half plane $\mathcal{H}$ by the action of Γ. One equips Y_Γ with the analytic structure coming from the projection map $\pi : \mathcal{H} \longrightarrow Y_\Gamma$. (More precisely, if $y = \pi(\tau)$, and $G_\tau \subset \Gamma$ is the stabilizer of τ in Γ, then the local ring $\mathcal{O}_{Y_\Gamma, y}$ is identified with the local ring of germs of holomorphic functions at τ which are invariant under the action of G_τ.) This makes Y_Γ into a connected complex analytic manifold of dimension one, i.e., a Riemann surface. If Γ is $\Gamma_0(N)$ (resp. $\Gamma_1(N)$, or $\Gamma(N)$), we will also denote Y_Γ by $Y_0(N)$ (resp. $Y_1(N)$, or $Y(N)$). One compactifies Y_Γ by adjoining a finite set of *cusps* which correspond to orbits of $\mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}$ under Γ. Call X_Γ the corresponding compact Riemann surface. (For more details, notably on the definition of the analytic structure on X_Γ at the cusps, see for example [Kn], p. 311, or [Shi2], ch. 1.) It follows from the definition of this analytic structure that the field K_Γ of meromorphic functions on X_Γ is equal to the set of meromorphic functions on $\mathcal{H}$ satisfying

- (Transformation property): $f(\gamma\tau) = f(\tau)$, for all $\gamma \in \Gamma$;
- (Behaviour at the cusps): For all $\gamma \in SL_2(\mathbb{Z})$, the function $f(\gamma\tau)$ has a Puiseux series expansion $\sum_{-m}^{\infty} a_n q^{n/h}$ in fractional powers of $q = e^{2\pi i\tau}$.

Riemann's existence theorem (cf. for example [For], ch. 2) asserts that the analytic structure on X_Γ comes from an algebraic one, i.e., the field K_Γ is a finitely generated extension of $\mathbb{C}$ of transcendence degree 1. Thus we can, and will, view X_Γ as a complex algebraic curve over $\mathbb{C}$. If Γ is $\Gamma_0(N)$ (resp. $\Gamma_1(N)$, or $\Gamma(N)$), we will also denote X_Γ by $X_0(N)$ (resp. $X_1(N)$, or $X(N)$).

Examples and exercises:
1. For $N = 1$, the curve $X_0(N) = X_1(N) = X(N)$ is a curve of genus 0, and its field of functions is the ring $\mathbb{C}(j)$, where j is the classical modular function,

$$j(\tau) = q^{-1} + 744 + 196884q + \cdots, \quad q = e^{2\pi i\tau}.$$

(Cf., for example, [Se4], ch. 7.)

2. For $\tau \in \mathcal{H} \cup \mathbb{P}^1(\mathbb{Q}) = \bar{\mathcal{H}}$, define G_τ to be the stabilizer of τ in $PSL_2(\mathbb{Z})$, and let $e_\tau = \#(G_\tau/(G_\tau \cap \Gamma))$. Show that e_τ depends only on the Γ-orbit of τ in $\bar{\mathcal{H}}$, and that $e_\tau = 1$ for all but finitely many τ in $\bar{\mathcal{H}}/\Gamma$. Using the Riemann-Hurwitz formula (cf. [Ki], sec. 4.3) show that the genus of X_Γ is given by

$$g(\Gamma) = 1 - [PSL_2(\mathbb{Z}) : \Gamma] + \frac{1}{2} \sum_{\tau \in \bar{\mathcal{H}}/\Gamma} (e_\tau - 1).$$

Use this to compute the genus of $X_0(p)$, $X_1(p)$, and $X(p)$ for p prime. For details, see [Shi2], sec. 1.6 or [Ogg].

3. For $\Gamma = \Gamma(2)$, show that X_Γ is isomorphic to $\mathbb{P}^1$, and that Y_Γ is isomorphic to $\mathbb{P}^1 - \{0, 1, \infty\}$. Show that $\Gamma/\langle \pm 1 \rangle$ is the free group on the two generators $g_1 = \left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right)$ and $g_2 = \left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right)$.

4. Define a homomorphism $\Gamma(2) \longrightarrow \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$, by sending g_1 to $(1, 0)$ and g_2 to $(0, 1)$, and let Γ denote its kernel. Show that Γ is not in general a congruence subgroup and that the curve $Y_\Gamma := \mathcal{H}/\Gamma$ is birationally isomorphic to the Fermat curve of degree n with affine equation $x^n + y^n = 1$.

Moduli interpretations: The points in $Y_\Gamma = \mathcal{H}/\Gamma$ can be interpreted as elliptic curves over $\mathbb{C}$ with some extra "level N" structure. More precisely,

- If $\Gamma = \Gamma_0(N)$, then the Γ-orbit of $\tau \in \mathcal{H}$ corresponds to the complex torus $E = \mathbb{C}/\langle 1, \tau \rangle$ with the distinguished cyclic subgroup of order N generated by $\frac{1}{N}$. Thus, points on $Y_0(N)$ parametrize isomorphism classes of pairs (E, C) where E is an elliptic curve over $\mathbb{C}$ and C is a cyclic subgroup of E of order N.

- If $\Gamma = \Gamma_1(N)$, then the Γ-orbit of τ corresponds to the complex torus $E = \mathbb{C}/\langle 1, \tau \rangle$ with the distinguished point of order N given by $\frac{1}{N}$. Hence, points on $Y_1(N)$ parametrize isomorphism classes of pairs (E, P) where now P is a point on E of exact order N.

Remark 1.10 One checks that the above rules set up a bijection between points on Y_Γ and elliptic curves with the appropriate structures, and that the projection $Y_1(N) \longrightarrow Y_0(N)$ sending $\Gamma_1(N)\tau$ to $\Gamma_0(N)\tau$ becomes the "forgetful" map sending (E, P) to $(E, \langle P \rangle)$.

Remark 1.11 (This remark will be used in section 1.3 when discussing Hecke operators.) Define an n-isogeny of Γ-structures to be an n-isogeny of the underlying elliptic curves which sends one Γ-structure to the other. If p is a prime not dividing N, then there are exactly $p+1$ distinct p-isogenies from $(\mathbb{C}/\langle \tau, 1 \rangle, \frac{1}{N})$, whose images are the pairs:

$$\left(\mathbb{C}/\left\langle \frac{\tau + i}{p}, 1 \right\rangle, \frac{1}{N}\right) \quad (i = 0, \ldots, p-1), \qquad \left(\mathbb{C}/\langle p\tau, 1 \rangle, \frac{p}{N}\right).$$

If p divides N, then there are only p distinct p-isogenies from $(\mathbb{C}/\langle \tau, 1 \rangle, \frac{1}{N})$, since $(\mathbb{C}/\langle p\tau, 1 \rangle, \frac{p}{N})$ is not a $\Gamma_1(N)$-structure (the point p/N not being of exact order N on the complex torus $\mathbb{C}/\langle p\tau, 1 \rangle$).

Modular forms: Let k be an even positive integer. A *modular form* of weight k on Γ is a holomorphic function f on $\mathcal{H}$ satisfying:

- (Transformation property): $f(\gamma\tau) = (c\tau + d)^k f(\tau)$, for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$.
- (Behaviour at the cusps): For all $\gamma \in PSL_2(\mathbb{Z})$, the function $(c\tau + d)^{-k} f(\gamma\tau)$ has a Puiseux series expansion $\sum_0^\infty a_n q^{n/h}$ in fractional powers of $q = e^{2\pi i \tau}$. We call $\sum a_n q^{n/h}$ the Fourier expansion of f at the cusp $\gamma^{-1}(i\infty)$.

A modular form which satisfies the stronger property that the constant coefficient of its Fourier expansion at each cusp vanishes is called a *cusp form.* We denote by $M_k(\Gamma)$ the complex vector space of modular forms of weight k on Γ, and by $S_k(\Gamma)$ the space of cusp forms on Γ. (For examples, see [DI], sec. 2.2 and the references therein, especially, [Shi2], ch. 2.)

This article is mainly concerned with modular forms of weight 2, and hence we will focus our attention on these from now on. A pleasant feature of the case $k = 2$ is that the cusp forms in $S_2(\Gamma)$ admit a direct geometric interpretation as holomorphic differentials on the curve X_Γ.

Lemma 1.12 *The map $f(\tau) \mapsto \omega_f := 2\pi i f(\tau) d\tau$ is an isomorphism between the space $S_2(\Gamma)$ and the space $\Omega^1(X_\Gamma)$ of holomorphic differentials on the curve X_Γ.*

Sketch of proof: One checks that the transformation property satisfied by $f(\tau)$ under Γ causes the expression $f(\tau)d\tau$ to be Γ-invariant, and that the condition of vanishing at the cusps translates into holomorphicity of $f(\tau)d\tau$. (Note, for example, that $2\pi i d\tau = dq/q$, so that ω_f is holomorphic at $i\infty$ precisely when $f(q)$ vanishes at $q = 0$.) □

As a corollary, we find:

Corollary 1.13 *The space $S_2(\Gamma)$ is finite-dimensional, and its dimension is equal to the genus g of X_Γ.*

Proof: This follows directly from the Riemann-Roch theorem, cf. [Ki], sec. 6.3. □

To narrow still further the focus of our interest, we will be mostly concerned with the cases $\Gamma = \Gamma_0(N)$ and $\Gamma_1(N)$. A slightly more general framework is sometimes convenient, so we suppose from now on that Γ satisfies

$$\Gamma_1(N) \subset \Gamma \subset \Gamma_0(N).$$

Such a group Γ is necessarily of the form $\Gamma_H(N)$ for some subgroup H of $(\mathbb{Z}/N\mathbb{Z})^\times$. Because the transformation $\tau \mapsto \tau + 1$ belongs to Γ the forms in $S_2(\Gamma)$ are periodic functions on $\mathcal{H}$ of period 1, and hence their Fourier expansions at $i\infty$ are of the form

$$f(\tau) = \sum_{n>0} a_n q^n, \quad q = e^{2\pi i \tau}, a_n \in \mathbb{C}.$$

The Petersson inner product: The spaces $S_2(\Gamma)$ are also equipped with a natural Hermitian inner product given by

$$\langle f, g \rangle = \frac{i}{8\pi^2} \int_{X_\Gamma} \omega_f \wedge \bar{\omega}_g = \int_{\mathcal{H}/\Gamma} f(\tau)\bar{g}(\tau) dx dy,$$

where $\tau = x + iy$. This is called the Petersson inner product.

The diamond operators: Suppose now that $\Gamma = \Gamma_1(N)$ and let d be an element of $(\mathbb{Z}/N\mathbb{Z})^\times$. The map $\langle d \rangle$ which sends an elliptic curve with Γ-structure (E, P) to the pair (E, dP) gives an automorphism of Y_Γ which extends to X_Γ. It is called the *diamond operator*. For τ in $\mathcal{H}$ and $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\Gamma_0(N)$, we have

$$\langle d \rangle(\Gamma\tau) = \Gamma(\gamma\tau).$$

Hence $\langle d \rangle$ acts on $S_2(\Gamma)$, identified with the holomorphic differentials on X_Γ, by the rule

$$\langle d \rangle f(\tau) = (c\tau + d)^{-2} f\left(\frac{a\tau + b}{c\tau + d}\right).$$

In geometric terms, the diamond operators are the Galois automorphisms of the natural (branched) covering $X_1(N) \longrightarrow X_0(N)$ whose Galois group is

isomorphic to $\Gamma_0(N)/\langle \pm\Gamma_1(N)\rangle = (\mathbb{Z}/N\mathbb{Z})^\times/\langle \pm 1\rangle$. Given an even Dirichlet character $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \longrightarrow \mathbb{C}^\times$, say that f is a modular form of level N and character χ if it belongs to the χ-eigenspace in $S_2(\Gamma_1(N))$ under this action. Let $S_2(N,\chi)$ denote the space of all such forms. Thus a function f in $S_2(N,\chi)$ is a cusp form on $\Gamma_1(N)$ which satisfies the stronger transformation property:

$$f\left(\frac{a\tau+b}{c\tau+d}\right) = \chi(d)(c\tau+d)^2 f(\tau), \text{ for all } \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Gamma_0(N).$$

Note that if $\mathbf{1}$ is the trivial character, then $S_2(N,\mathbf{1})$ is canonically identified with $S_2(\Gamma_0(N))$, which we will also denote by $S_2(N)$. Finally note the direct sum decomposition:

$$S_2(\Gamma_1(N)) = \oplus_\chi S_2(N,\chi),$$

where the sum ranges over all the even Dirichlet characters modulo N.

Exercise 1.14 Show that if $f(\tau)$ belongs to $S_2(N)$, then $f(a\tau)$ belongs to $S_2(mN)$, for each integer a dividing m.

Jacobians of modular curves: Let V be the dual space

$$V = S_2(\Gamma)^\vee := \mathrm{Hom}(S_2(\Gamma),\mathbb{C}).$$

It is a complex vector space of dimension $g = \mathrm{genus}(X_\Gamma)$. The integral homology $\Lambda = H_1(X_\Gamma,\mathbb{Z})$ maps naturally to V by sending a homology cycle c to the functional ϕ_c defined by $\phi_c(f) = \int_c \omega_f$. The image of Λ is a lattice in V, i.e., a $\mathbb{Z}$-module of rank $2g$ which is discrete (cf. [Mu1], cor. 3.8). Fix a base point $\tau_0 \in \mathcal{H}$, and define the Abel-Jacobi map $\Phi_{\mathrm{AJ}} : X_\Gamma(\mathbb{C}) \longrightarrow V/\Lambda$ by $\Phi_{\mathrm{AJ}}(P)(f) = \int_{\tau_0}^P \omega_f$. Note that this is well-defined, i.e., it does not depend on the choice of path on X_Γ from τ_0 to P, up to elements in Λ.

We extend the map Φ_{AJ} by linearity to the group $\mathrm{Div}\,(X_\Gamma)$ of divisors on X_Γ, and observe that the restriction of Φ_{AJ} to the group $\mathrm{Div}^0(X_\Gamma)$ of degree 0 divisors does not depend on the choice of base-point τ_0. Moreover we have the Abel-Jacobi theorem:

Theorem 1.15 *The map*

$$\Phi_{\mathrm{AJ}} : \mathrm{Div}^0(X_\Gamma) \longrightarrow V/\Lambda$$

has a kernel consisting precisely of the group $P(X_\Gamma)$ of principal divisors on X_Γ. Hence Φ_{AJ} induces an isomorphism from $\mathrm{Pic}^0(X_\Gamma) := \mathrm{Div}^0(X_\Gamma)/P(X_\Gamma)$ *to V/Λ.*

For the proof, see [Mu1], ch. 3. The quotient V/Λ is a complex torus, and is equal to the group of complex points of an abelian variety. We denote this abelian variety by J_Γ, the *Jacobian variety* of X_Γ over $\mathbb{C}$. If $\Gamma = \Gamma_0(N)$ or $\Gamma_1(N)$, we will also write $J_0(N)$ or $J_1(N)$ respectively for the Jacobian J_Γ.

1.3 Hecke operators and Hecke theory

We maintain our running assumption that Γ satisfies

$$\Gamma_1(N) \subset \Gamma \subset \Gamma_0(N).$$

If p is a prime not dividing the level N, we define the Hecke operator T_p on $S_2(\Gamma)$ by the formula

$$T_p(f) = \frac{1}{p}\sum_{i=0}^{p-1} f\left(\frac{\tau+i}{p}\right) + p\langle p\rangle f(p\tau).$$

We give a more conceptual description of T_p in terms of remark 1.11 in the case $\Gamma = \Gamma_1(N)$. We have

$$\omega_{T_p(f)} = \sum \phi_i^*(\omega_f),$$

where $\phi_i(\tau) = \frac{\tau+i}{p}$, and $\phi_\infty(\tau) = \langle p\rangle p\tau$ represent the $p+1$ curves with Γ-structure that are images of $(\mathbb{C}/\langle\tau,1\rangle, \frac{1}{N})$ by a p-isogeny, and the ϕ_i^* are the pull-back maps on differential forms on $\mathcal{H}$. (An isogeny of elliptic curves with Γ-structure is simply an isogeny between the underlying curves which sends one Γ-structure to the other.) Such a description makes it evident that $T_p(f)$ belongs to $S_2(\Gamma)$, if f does. In terms of the Fourier expansion of $f = \sum a_n q^n$, the formula for the operator T_p on $S_2(N,\chi)$ is given by:

$$T_p(f) = \sum_{p|n} a_n q^{n/p} + p\chi(p)\sum a_n q^{pn}.$$

If p divides N, then we define the Hecke operator U_p analogously, by summing again over all the cyclic p-isogenies of Γ-structures. Since there are only p of them, the formula becomes simpler:

$$U_p(f) = \frac{1}{p}\sum_{i=0}^{p-1} f\left(\frac{\tau+i}{p}\right) = \sum_{p|n} a_n q^{n/p}.$$

The reader is invited to check that the Hecke operators of the form T_p or U_q commute with each other, and also that they commute with the diamond operators introduced in the previous section.

We extend the definition of the Hecke operators to operators T_{p^n}, with $n > 1$, by the inductive formulae

$$T_{p^{n+1}} = T_p T_{p^n} - \langle p\rangle p T_{p^{n-1}}, \quad \text{if } (p,N) = 1,$$

and $T_{p^n} = U_p^n$ otherwise. We then define the operator T_n, where $n = \prod p_i^{e_i}$ is written as a product of powers of distinct primes p_i, by

$$T_n = \prod_i T_{p_i^{e_i}}.$$

This definition makes the Hecke operators multiplicative, i.e., $T_m T_n = T_{mn}$ if $(m,n) = 1$. (A more conceptual definition of the Hecke operator T_n is that $T_n(f)$ is obtained by summing the pullback of ω_f over the maps describing all the cyclic n-isogenies of Γ-structures.) The relations among the different Hecke operators can be stated succinctly by saying that they obey the following (formal) identity:

$$\prod_{p \not| N} (1 - T_p p^{-s} + \langle p \rangle p^{1-2s})^{-1} \prod_{p|N} (1 - U_p p^{-s})^{-1} = \sum_n T_n n^{-s}. \tag{1.3.1}$$

The reader can consult [DI], sec. 3 and the references therein (especially [Shi2], ch. 3 or [Kn]) for more details and different points of view on Hecke operators. Let $\mathbb{T}$ be the subring of $\mathrm{End}_{\mathbb{C}}(S_2(\Gamma))$ generated over $\mathbb{C}$ by all the Hecke operators T_p for $p \not| N$, U_q for $q|N$, and $\langle d \rangle$ acting on $S_2(\Gamma)$.

Definition 1.16 A modular form f is an *eigenform* if it is a simultaneous eigenvector for *all* the Hecke operators in $\mathbb{T}$, i.e., if there exists a $\mathbb{C}$-algebra homomorphism $\lambda : \mathbb{T} \longrightarrow \mathbb{C}$ such that $Tf = \lambda(T)f$, for all $T \in \mathbb{T}$.

A direct calculation shows that the coefficients a_n of an eigenform f can be recovered from the homomorphism λ by the formula:

$$a_n(f) = a_1(f)\lambda(T_n).$$

It follows that the first Fourier coefficient a_1 of a non-zero eigenform is always non-zero, and that the non-trivial eigenspaces for $\mathbb{T}$ are all one-dimensional:

Proposition 1.17 *Given a non-zero algebra homomorphism $\lambda : \mathbb{T} \longrightarrow \mathbb{C}$, there is exactly one eigenform f up to scaling, which satisfies $Tf = \lambda(T)f$, for all $T \in \mathbb{T}$.*

Sketch of proof: The proof of the existence of f is an exercise in commutative algebra (localize $S_2(\Gamma)$ at the kernel of λ), and the uniqueness is clear from the formula above. □

We call an eigenform satisfying $a_1 = 1$ a *normalized* eigenform.

Atkin-Lehner theory: It is natural to ask whether $S_2(\Gamma)$ can be decomposed into a basis consisting of distinct normalized eigenforms. Unfortunately, this is not always possible, as the following exercise illustrates.

Exercise 1.18 Suppose that p^3 divides N exactly. Let $\mathbb{T}'$ be the algebra of Hecke operators (generated by the operators T_q with $q \not| N/p^3$, and U_q with $q|N/p^3$) acting on $S_2(N/p^3)$. Let $f = \sum_{n=1}^{\infty} a_n q^n$ be a $\mathbb{T}'$-eigenform of level N/p^3 in $S_2(N/p^3)$. Show that the space S_f spanned by the forms $f(\tau)$, $f(p\tau)$, $f(p^2\tau)$, and $f(p^3\tau)$ is contained in $S_2(N)$, and is stable for the action of the Hecke operators T_q, $q \not| N$, and U_q, $q|N$. Show that S_f has no basis of simultaneous eigenforms for the Hecke algebra $\mathbb{T}$ of level N, so that the action of $\mathbb{T}$ on S_f is not semi-simple.

Let $\mathbb{T}^0$ denote the subalgebra of $\mathbb{T}$ generated only by the "good" Hecke operators T_q with $q \not| N$, and $\langle d \rangle$.

Proposition 1.19 *If q does not divide N, the adjoint of the Hecke operator T_q with respect to the Petersson scalar product is the operator $\langle q \rangle^{-1} T_q$, and the adjoint of $\langle q \rangle$ is $\langle q \rangle^{-1}$. In particular, the Hecke operators commute with their adjoints.*

Proof: See [Kn], th. 9.18 and 8.22, or [Ogg]. □

Proposition 1.19 implies, by the spectral theorem for commuting operators that commute with their adjoints:

Proposition 1.20 *The algebra $\mathbb{T}^0$ is semi-simple (i.e, it is isomorphic to a product $\mathbb{C} \times \cdots \times \mathbb{C}$ of a certain number of copies of $\mathbb{C}$), and there is a basis of $S_2(\Gamma)$ consisting of simultaneous eigenvectors for the operators T_q.*

Thus, $\mathbb{T}^0$ has the merit of being semi-simple, while $\mathbb{T}$ is not in general. The cost of replacing $\mathbb{T}$ by $\mathbb{T}^0$, however, is that one loses "multiplicity one", i.e., the eigenspaces for $\mathbb{T}^0$ need not be one-dimensional. For example, the space S_f defined in the previous exercise is contained in a single eigenspace for $\mathbb{T}^0$.

The theory of Atkin-Lehner [AL] gives essentially a complete understanding of the structure of the algebra $\mathbb{T}$, and the structure of the space of eigenforms. To motivate the main result, observe that the problem in the exercise above seems to be caused by forms of level N that are coming from forms of lower level N/p^3 by a straightforward operation, and are therefore not "genuinely" of level N. They are the analogues, in the context of modular forms, of non-primitive Dirichlet characters.

Definition 1.21 We define the *old* subspace of $S_2(\Gamma)$ to be the space spanned by those functions which are of the form $g(az)$, where g is in $S_2(\Gamma_1(M))$ for some $M < N$ and aM divides N. We define the *new* subspace of $S_2(\Gamma)$ to be the orthogonal complement of the old subspace with respect to the Petersson scalar product. A normalized eigenform in the new subspace is called a *newform of level N*.

The following result is the main consequence of the theory of Atkin-Lehner. It gives a complete answer to the question of what is the structure of the algebra $\mathbb{T}$ acting on $S_2(\Gamma)$.

Theorem 1.22 *If f is in the new subspace of $S_2(\Gamma)$ and is an eigenvector for all the operators in $\mathbb{T}^0$, then it is also an eigenform for $\mathbb{T}$, and hence is unique up to scaling. More generally, if f is a newform of level $N_f | N$, then the space S_f defined by*

$$S_f = \{g \in S_2(\Gamma) \text{ such that } Tg = \lambda_f(T)g, \text{ for all } T \in \mathbb{T}^0\}$$

is stable under the action of all the Hecke operators in $\mathbb{T}$. *It is spanned by the linearly independent forms* $f(az)$ *where* a *ranges over the divisors of* N/N_f. *Furthermore, we have*

$$S_2(\Gamma) = \oplus_f S_f,$$

where the sum is taken over all newforms f *of some level* N_f *dividing* N.

See [AL] for the proof in the case $\Gamma = \Gamma_0(N)$, and [La2], ch. VIII for the general case. (See also [DI], sec. 6 for an overview.)

Exercise 1.23 Consider the case where $\Gamma = \Gamma_0(22)$. Show that $X_0(22)$ is of genus 2, and hence that $S_2(22)$ has dimension 2. Show that $S_2(22)$ is equal to S_f, where $f = (\eta(\tau)\eta(11\tau))^2$ is a newform of level 11, so that in particular there are no newforms of level 22 on Γ. Show that $\mathbb{T}^0$ is isomorphic to $\mathbb{C}$ in this case, and that $\mathbb{T}$ is isomorphic to the semisimple algebra $\mathbb{C} \times \mathbb{C}$.

Action on homology and Jacobians: Note that the Hecke operators act on $V = S_2(\Gamma)^\vee$ by duality. One checks (cf. [Kn], props. 11.23, 11.24) that the sublattice Λ of V is stable under the action of all the Hecke operators T_n, and of the diamond operators $\langle d \rangle$. Therefore the operators T_n and $\langle d \rangle$ give rise to endomorphisms of the torus V/Λ, and hence the Jacobian variety J_Γ, in a natural way. The involution $\tau \mapsto -\bar{\tau}$ gives rise to an involution on $X_\Gamma(\mathbb{C})$ (which is complex conjugation on the model of X_Γ over $\mathbb{R}$ deduced from the $\mathbb{Q}$- model defined in section 1.5). Since complex conjugation is continuous it also acts on the integral homology $\Lambda = H_1(X_\Gamma(\mathbb{C}), \mathbb{Z})$. Let Λ^+ and Λ^- be the sublattices of Λ on which complex conjugation acts by $+1$ and -1. These are sublattices of Λ of rank g which are stable under the Hecke operators, since complex conjugation commutes with the Hecke action.

A more algebraic description of the action of T_p on J_Γ is given via the notion of an *algebraic correspondence*. A correspondence on a curve X is a divisor C on $X \times X$ which is taken modulo divisors of the form $\{P\} \times X$ and $X \times \{Q\}$. Let π_1 and π_2 denote the projections of $X \times X$ onto each factor. Then the correspondence C induces a map on divisors of X, by setting

$$C(D) = \pi_2(\pi_1^{-1}(D) \cdot C).$$

(For the definition of the intersection $D_1 \cdot D_2$ of two divisors, see [We1].) The map C preserves divisors of degree 0, and sends principal divisors on X to principal divisors. It gives a well defined algebraic endomorphism of the Jacobian variety $\mathrm{Jac}(X)$. Given a correspondence C, its transpose $C^\vee$ is defined to be the divisor of $X \times X$ obtained by interchanging the two factors of $X \times X$. One can define a natural notion of composition for correspondences, and the set of correspondences forms a ring. The general theory of correspondences and the proofs of the above facts are given in [We1], particularly the second chapter.

The Hecke correspondence T_n is defined as the closure in $X_\Gamma \times X_\Gamma$ of the locus of points (A, B) in $Y_\Gamma \times Y_\Gamma$ such that there is a degree n cyclic isogeny

of elliptic curves with Γ-structure from A to B. For example, if p is a prime not dividing N, then T_p is an algebraic curve in $X_1(N) \times X_1(N)$ which is birational to $X_{\Gamma_1(N)\cap\Gamma_0(p)}$. The induced map on divisors in this case satisfies

$$T_p((E,P)) = \sum (E/C, P \bmod C)$$

where the sum runs over the subgroups C of E having order p. Note also that if (A,B) belongs to T_p, then the isogeny dual to $A \longrightarrow B$ gives a p-isogeny from B to $\langle p\rangle A$, so that

$$T_p^\vee = \langle p\rangle^{-1} T_p.$$

1.4 The L-function associated to a cusp form

For this section, let f in $S_2(\Gamma_1(N))$ be a cusp form with Fourier expansion at $i\infty$ given by $\sum_n a_n q^n$. One has the following estimate for the size of the Fourier coefficients a_n:

Theorem 1.24 *The coefficients $a_n \in \mathbb{C}$ satisfy the inequality*

$$|a_n| \leq c(f)\sigma_0(n)\sqrt{n},$$

where $c(f)$ is a constant depending only on f, and $\sigma_0(n)$ denotes the number of positive divisors of n.

Sketch of proof: This follows from proposition 1.51 of section 1.7 which relates the p-th Fourier coefficients of eigenforms, for p a prime not dividing the level of Γ, to the number of points on certain abelian varieties over the finite field $\mathbb{F}_p$. The estimates of Hasse and Weil for the number of points on abelian varieties over finite fields (stated in theorem 1.4 of section 1.1 for the special case of elliptic curves; see [Wel], §IV for the general case) thus translate into asymptotic bounds for the Fourier coefficients of these eigenforms. We note that the cruder estimate $|a_n| = O(n)$, which is enough for the purposes of this section, can be derived by a more elementary, purely analytic argument; cf. [Ogg], ch. IV, prop. 16. □

The L-function associated to f is defined by the formula:

$$L(f,s) = \sum a_n n^{-s}.$$

As in exercise 1.8, one can show that the infinite sum defining $L(f,s)$ converges absolutely in the right half-plane $\mathrm{Re}\,(s) > \frac{3}{2}$. A much better insight is gained into the function $L(f,s)$ by noting that it is essentially the Mellin transform of the modular form f. More precisely, if we set $\Lambda(f,s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f,s)$, then we have

$$\Lambda(f,s) = N^{s/2}\int_0^\infty f(iy)y^s dy/y \qquad (1.4.1)$$

Exercise 1.25 Check the formula above.

This integral representation for $L(f,s)$ gives the analytic continuation of $L(f,s)$ to the entire complex plane. The modular invariance of f translates into a functional equation for $L(f,s)$: more precisely, let w_N be the Atkin-Lehner involution defined by $w_N(\tau) = -1/N\tau$. The reader may check that w_N induces an involution of X_Γ, and hence of $\Omega^1(X_\Gamma) = S_2(\Gamma)$. One finds that $L(f,s)$ satisfies the functional equation:

$$\Lambda(f,s) = -\Lambda(w_N(f), 2-s).$$

For a proof of this, see [Ogg], ch. V, lemma 1. Eigenforms for $\mathbb{T}$ in $S_2(N,\chi)$ have a great importance in the theory because their associated L-functions have an Euler product expansion, in addition to an analytic continuation and functional equation:

Theorem 1.26 *If $f = \sum a_n q^n$ is a normalized eigenform in $S_2(N,\chi)$ for all the Hecke operators, then the L-function $L(f,s) = \sum a_n n^{-s}$ has the Euler product expansion*

$$\prod_{p \not| N} (1 - a_p p^{-s} + \chi(p) p^{1-2s})^{-1} \prod_{p|N} (1 - a_p p^{-s})^{-1}.$$

Proof: This follows directly from equation (1.3.1) of section 1.3. □

If f is a newform of level N, then it is also an eigenform for w_N, so that the functional equation may be viewed as relating $L(f,s)$ and $L(f,2-s)$. We can also state the following more precise version of theorem 1.24 (see lemma 3.2 of [Hi2] for example for parts (b), (c) and (d)).

Theorem 1.27 *Suppose that f is a newform of level N_f and let N_χ denote the conductor of its character χ.*

(a) *If p does not divide N_f then $|a_p| \leq 2\sqrt{p}$.*

(b) *If $p||N_f$ and p does not divide N_χ then $a_p^2 = \chi_0(p)$ where χ_0 is the primitive character associated to χ.*

(c) *If p divides N_f and p does not divide N_f/N_χ then $|a_p| = \sqrt{p}$.*

(d) *If p^2 divides N_f and p divides N_f/N_χ then $a_p = 0$.*

1.5 Modular curves and modular forms over $\mathbb{Q}$

Modular curves: For Γ between $\Gamma_0(N)$ and $\Gamma_1(N)$, the modular curve X_Γ has a model over $\mathbb{Q}$. We describe such a model in the case of $\Gamma = \Gamma_0(N)$; the construction for general Γ follows from similar considerations.

The key remark here is that, as was noted in section 1.2, the complex points on the curve $Y_0(N)$ have a natural interpretation as moduli of elliptic

curves together with a cyclic subgroup of order N, given by sending the point $\tau \in \mathcal{H}/\Gamma_0(N)$ to the pair $(\mathbb{C}/\langle 1, \tau\rangle, \langle 1/N\rangle)$.

Consider the "universal elliptic curve"

$$E_j : y^2 + xy = x^3 - \frac{36}{j - 1728}x - \frac{1}{j - 1728}.$$

It is an elliptic curve over the function field $\mathbb{Q}(j)$, with j-invariant j. Let d be the order of $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$, and let $C_1, \ldots, C_d$ denote the set of all cyclic subgroups of E_j of order N, defined over $\overline{\mathbb{Q}(j)}$, an algebraic closure of $\mathbb{Q}(j)$. Fix one of these subgroups, C. The Galois group $\mathrm{Gal}(\overline{\mathbb{Q}(j)}/\mathbb{Q}(j))$ permutes the C_i in a natural way. Let F_N be the smallest extension of $\mathbb{Q}(j)$ (viewed as embedded in $\overline{\mathbb{Q}(j)}$) with the property that $\sigma(C) = C$, for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}(j)}/F_N)$. It can be seen (cf. [Shi2], thm. 6.6) that the Galois action on the C_i is transitive so that $F_N/\mathbb{Q}(j)$ is of degree d, and that it is a *regular* extension, i.e., $F_N \cap \bar{\mathbb{Q}} = \mathbb{Q}$. Geometrically, F_N can be viewed as the function field of a curve $X_{/\mathbb{Q}}$ over $\mathbb{Q}$, with the inclusion of $\mathbb{Q}(j)$ corresponding to a map from $X_{/\mathbb{Q}}$ to the projective "j-line" over $\mathbb{Q}$. The pair (E_j, C) is an elliptic curve over $\mathbb{Q}(j)$ with a subgroup of order N defined over F_N. Using (E_j, C), each complex point of X gives an elliptic curve over $\mathbb{C}$ with a subgroup of order N, provided the point does not lie over $j = 0$, 1728 or ∞. The resulting map to $X_0(N)$ extends to an isomorphism from X to $X_0(N)$. The curve X thus constructed, together with this identification, is the desired model of $X_0(N)$ over $\mathbb{Q}$.

More concretely, the functions $j = j(\tau)$ and $j_N = j(N\tau)$ are related by a polynomial equation $\Phi_N(j, j_N) = 0$ with coefficients in $\mathbb{Q}$, of bidegree d. The field F_N is the function field of the affine curve $\Phi_N(X, Y) = 0$, and the mapping $\tau \mapsto (j(\tau), j(N\tau))$ gives a birational equivalence between $\mathcal{H}/\Gamma_0(N)$ and the complex curve defined by the equation Φ_N. In practice it is not feasible to write down the polynomial Φ_N, except for certain very small values of N. To study the models over $\mathbb{Q}$ of $X_0(N)$, more indirect methods are needed, which rely crucially on the moduli interpretation of $X_0(N)$. Similar remarks hold for $X_1(N)$.

Models over $\mathbb{Z}$: The work of Igusa [Ig], Deligne-Rapoport [DR], Drinfeld [Dr], and Katz-Mazur [KM] uses the moduli-theoretic interpretation to describe a canonical proper model for X_Γ over Spec $\mathbb{Z}$. These models allow us to talk about the reduction of X_Γ over finite fields $\mathbb{F}_p$, for p prime. The curve has good reduction at primes p not dividing N, with the "non-cuspidal" points of $X_{\Gamma/\mathbb{F}_p}$ corresponding to elliptic curves over $\bar{\mathbb{F}}_p$ with Γ-structure. The singular fibers at primes p dividing N can also be described precisely; an important special case (see [DR]) is that of $\Gamma_0(N)$ with p exactly dividing N. For further discussion and references, see [DI], sec. 8, 9.

From now on, when we write X_Γ, $X_0(N)$, or $X_1(N)$, we will mean the curve over $\mathbb{Q}$ which are the models described above for the complex curves defined in section 1.2.

Remark 1.28 When considering q-expansions, it is more convenient to use a different set of models over $\mathbb{Z}$ for these complex curves. We define $X_1^\mu(N)$ in the case of $\Gamma_1(N)$ as a model over $\mathbb{Z}$ which parametrizes pairs (E, i) where i is an embedding of μ_N in the (generalized) elliptic curve E. (So $X_1^\mu(N)$ is the model denoted $X_\mu(N)$ in [DI], sec. 9.3, assuming $N > 4$.) For Γ between $\Gamma_0(N)$ and $\Gamma_1(N)$, we define X_Γ^μ as the corresponding quotient of $X_1^\mu(N)$. This model has good reduction at primes p not dividing N, but unlike the models mentioned above, its fibers at primes p dividing N are smooth and irreducible, but not proper. In the case of $\Gamma = \Gamma_0(N)$, the curve $X_{\Gamma,\mathbb{Q}}^\mu$ can be identified with $X_0(N)$. However, this is not the case in general: the cusp ∞ is a rational point of $X_{\Gamma,\mathbb{Q}}^\mu$ but not necessarily of X_Γ.

Jacobians: Weil's theory [Wel] of the Jacobian shows that the Jacobians J_Γ defined in section 1.2 as complex tori also admit models over $\mathbb{Q}$. When we speak of J_Γ, $J_0(N)$ and $J_1(N)$ from now on, we will refer to these as abelian varieties defined over $\mathbb{Q}$. Thus, the points in $J_\Gamma(K)$, for any $\mathbb{Q}$-algebra K, are identified with the divisor classes on X_Γ of degree 0, defined over K.

We let $J_{\Gamma/\mathbb{Z}}$, denote the Néron model of the Jacobian J_Γ over $\mathrm{Spec}\,(\mathbb{Z})$. Using this model we define $J_{\Gamma/A}$ for arbitrary rings A. In particular we can consider $J_{\Gamma/\mathbb{F}_p}$, the reduction of the Jacobian in characteristic p, which is closely related to the reduction of the integral model of the curve X_Γ mentioned above. In particular, if p does not divide the level of Γ, then $J_{\Gamma/\mathbb{F}_p}$ can be identified with the Jacobian of $X_{\Gamma/\mathbb{F}_p}$. For a treatment of the case $\Gamma = \Gamma_0(N)$ with p exactly dividing N, see the appendix of [Maz1]; for more general discussion and references, see [DI], sec. 10, especially sec. 10.3.

Hecke operators: The Hecke operators have a natural moduli interpretation, which was already touched upon in section 1.3. In particular, one finds that the operator T_n arises from a correspondence which is defined over $\mathbb{Q}$, and gives rise to an endomorphism of the Jacobian J_Γ which is defined over $\mathbb{Q}$. This in turn gives rise to an endomorphism of the Néron model $J_{\Gamma/\mathbb{Z}}$, and we can then consider the endomorphism T_n on the reduction of the Jacobian in characteristic p. Recall that if p is a prime not dividing N, we may identify this reduction with the Jacobian of $X_{\Gamma/\mathbb{F}_p}$. (Cf. [MW], ch. 2, sec. 1, prop. 2.) Furthermore, one can show that the moduli-theoretic interpretation of the Hecke operator remains valid in characteristic p; i.e., the endomorphism T_n of $J_{\Gamma/\mathbb{F}_p}$ is induced by a map on divisors satisfying, for all ordinary elliptic curves A with Γ-structure:

$$T_n(A) = \sum_i i(A),$$

where the sum is taken over all cyclic isogenies of degree n. (See [DI], sec. 10.2 for further discussion and references.)

This description allows one to analyse, for example, the Hecke operator T_p over $\mathbb{F}_p$, when $(p, N) = 1$. Let us work with $\Gamma = \Gamma_1(N)$, to illustrate the idea.

For a variety X over $\bar{\mathbb{F}}_p$, let ϕ_X be the Frobenius morphism on X defined by raising coordinates to the pth power. Thus if (E,P) corresponds to a point of $X_1(N)_{/\mathbb{F}_p}$, then ϕ_E is an isogeny of degree p from (E,P) to the pair $(E_\infty,P_\infty)=\phi_{X_1(N)}(E,P)$. The graph of $\phi_{X_1(N)}$ in $(X_1(N)\times X_1(N))_{/\mathbb{F}_p}$ is a correspondence of degree p, which we call F. Let F' be the transpose of this correspondence. The endomorphism F of J_Γ induced by F is the Frobenius endomorphism ϕ_{J_Γ}, and the endomorphism F' is the dual endomorphism (in the sense of duality of abelian varieties). Now consider the divisor

$$F'((E,P))=(E_1,P_1)+\cdots+(E_p,P_p),$$

where the (E_i,P_i) are elliptic curves with Γ-structure in characteristic p. Since ϕ_{E_i} is an isogeny of degree p from (E_i,P_i) to (E,P), we also have the dual isogeny from (E,P) to (E_i,pP_i). If E is ordinary at p, then the points (E_∞,P_∞), (E_1,pP_1), $\ldots$, (E_p,pP_p) are a complete list of the distinct curves with Γ-structure which are p-isogenous to (E,P). Hence one has the equality of divisors on $X_1(N)_{/\mathbb{F}_p}$:

$$T_p((E,P))=(E_\infty,P_\infty)+(E_1,pP_1)+\cdots+(E_p,pP_p)=(F+\langle p\rangle F')((E,P)).$$

Since the ordinary points are dense on $X_1(N)_{/\mathbb{F}_p}$, we deduce that $T_p=(F+\langle p\rangle F')$ as endomorphisms of $J_1(N)_{/\mathbb{F}_p}$. This equation, known as Eichler-Shimura congruence relation, plays a central role in the theory. (For more details, see [DI], sec. 10.2, 10.3.)

Theorem 1.29 *If $p\nmid N$ then the endomorphism T_p of $J_{\Gamma/\mathbb{F}_p}$ satisfies*

$$T_p=F+\langle p\rangle F'.$$

Remark 1.30 This was proved by Eichler [Ei] to hold for all but finitely many p in the case of $\Gamma_0(N)$, and by Shimura ([Shi1], see also [Shi2], ch. 7) in the case of $\Gamma_1(N)$. The fact that it holds for all p not dividing N follows from work of Igusa [Ig].

Modular forms: In the same way that modular curves have models over $\mathbb{Q}$ and over $\mathbb{Z}$, the Fourier coefficients of modular forms also have natural rationality and integrality properties. We start by sketching the proof of:

Theorem 1.31 *The space $S_2(\Gamma)$ has a basis consisting of modular forms with integer Fourier coefficients.*

Proof: The Hecke operators act on the integral homology Λ^+ in a way that is compatible with the action on $S_2(\Gamma)$ and respects the natural (Poincaré) duality between these two spaces. Hence, if $\{\lambda_n\}_{n\in\mathbb{N}}$ is a system of eigenvalues for the T_n, then the λ_n are algebraic integers in some finite extension K of $\mathbb{Q}$, and the system $\{\lambda_n^\sigma\}_{n\in\mathbb{N}}$ is a system of eigenvalues for the T_n for any Galois automorphism σ of $\bar{\mathbb{Q}}/\mathbb{Q}$. Hence, we have shown:

Proposition 1.32 *If $f \in S_2(M, \chi)$ is a newform of some level M dividing N, then its Fourier coefficients lie in a finite extension K of $\mathbb{Q}$. Moreover, if $\sigma \in \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ is any Galois automorphism, then the Fourier series f^σ obtained by applying σ to the Fourier coefficients is a newform in $S_2(M, \chi^\sigma)$.*

The explicit description of $S_2(\Gamma)$ given in section 1.3 implies that $S_2(\Gamma)$ is spanned by forms having Fourier coefficients which are algebraic integers in some finite (Galois) extension K of $\mathbb{Q}$, and that the space of forms with Fourier coefficients in K is stable under the natural action of $\mathrm{Gal}(K/\mathbb{Q})$ on Fourier expansions. An application of Hilbert's theorem 90 shows that $S_2(\Gamma)$ has a basis consisting of forms with rational Fourier expansions, and the integrality of the Fourier coefficients of eigenforms yields the integrality statement of theorem 1.31. □

We define $S_2(\Gamma, \mathbb{Z})$ to be the space of modular forms with integral Fourier coefficients in $S_2(\Gamma)$. Theorem 1.31 states that $S_2(\Gamma, \mathbb{Z}) \otimes \mathbb{C} = S_2(\Gamma)$. Given any ring A, we define

$$S_2(\Gamma, A) = S_2(\Gamma, \mathbb{Z}) \otimes A,$$

and define $S_2(N, A)$ and $S_2(N, \chi, A)$ (where χ now is a character with values in $A^\times$) in the obvious way. If A is contained in $\mathbb{C}$, the q-expansion principle below allows us to identify $S_2(\Gamma, A)$ with the set of modular forms in $S_2(\Gamma)$ with Fourier coefficients in A.

The q-expansion principle: Because the modular curve $X_0(N)$ has a model over $\mathbb{Q}$, the space of modular forms $S_2(N) = \Omega^1(X_0(N))$ has a natural rational structure, given by considering the differential forms on $X_0(N)$ defined over $\mathbb{Q}$. The fundamental q-expansion principle (see [DR], ch. 7, or [Kat], sec. 1.6) says that these algebraic structures are the same as those obtained analytically by considering q-expansions at ∞. More generally, using the model X_Γ^μ, we obtain the q-expansion principle over $\mathbb{Z}$ for cusp forms on Γ (cf. [DI], sec. 12.3).

Theorem 1.33 *The map $S_2(\Gamma) \longrightarrow \Omega^1(X_\Gamma)$ defined by $f \mapsto \omega_f$ induces an isomorphism from $S_2(\Gamma, \mathbb{Z})$ to $\Omega^1(X_\Gamma^\mu)$. Furthermore, if A is flat over $\mathbb{Z}$ or N is invertible in A, then the induced map $S_2(\Gamma, A) \to \Omega^1(X_{\Gamma,A}^\mu)$ is an isomorphism. Furthermore, if A is a subring of $\mathbb{C}$, then this isomorphism identifies $S_2(\Gamma, A)$ with set of modular forms in $S_2(\Gamma)$ having coefficients in A.*

1.6 The Hecke algebra

It follows directly from the formulas for the Hecke operators acting on q-expansions that the T_n leave $S_2(\Gamma_0(N), \mathbb{Z})$ stable, as well as the subspace of $S_2(N, \chi)$ with coefficients in $\mathbb{Z}[\chi]$. Using the q-expansion principle (theorem 1.33), one can also show ([DI], sec. 12.4) that the diamond operators

preserve the spaces of cusp forms on Γ with integral Fourier expansions, and hence that the space $S_2(\Gamma, \mathbb{Z})$ is preserved by all the Hecke operators.

We define $\mathbb{T}_{\mathbf{Z}}$ to be the ring generated over $\mathbb{Z}$ by the Hecke operators T_n and $\langle d \rangle$ acting on the space $S_2(\Gamma, \mathbb{Z})$. More generally, if A is any ring, we define $\mathbb{T}_A$ to be the A-algebra $\mathbb{T}_{\mathbf{Z}} \otimes A$. This Hecke ring acts on the space $S_2(\Gamma, A)$ in a natural way. Before studying the structure of the Hecke rings $\mathbb{T}_A$ as we vary the rings A, we note the following general result (Cf. [Shi2], ch. 3.):

Lemma 1.34 *The space* $S_2(\Gamma, A)^\vee = \mathrm{Hom}_A(S_2(\Gamma, A), A)$ *is a free* $\mathbb{T}_A$*-module of rank one.*

Sketch of proof: One checks that the pairing $\mathbb{T}_{\mathbf{Z}} \times S_2(\Gamma, \mathbb{Z}) \longrightarrow \mathbb{Z}$ defined by $(T, f) \mapsto a_1(Tf)$ sets up a perfect, $\mathbb{T}_{\mathbf{Z}}$-equivariant duality between $\mathbb{T}_{\mathbf{Z}}$ and $S_2(\Gamma, \mathbb{Z})$. The result for arbitrary A follows. □

Hecke rings over $\mathbb{C}$: If $A = \mathbb{C}$, then the structure of the ring $\mathbb{T} = \mathbb{T}_{\mathbf{C}}$ is completely described by theorem 1.22. More precisely, if $\mathbb{T}_{\mathbf{C},f}$ denotes the image of the Hecke algebra acting on the space S_f defined in section 1.3, then

$$\mathbb{T}_{\mathbf{C}} = \oplus_f \mathbb{T}_{\mathbf{C},f},$$

where the direct sum ranges over all distinct newforms f of some level N_f dividing N. Furthermore, the algebra $\mathbb{T}_{\mathbf{C},f}$ can be described explicitly. In particular, if f is a newform of level N then $\mathbb{T}_{\mathbf{C},f}$ is isomorphic to $\mathbb{C}$, but if N_f is not equal to N then the ring $\mathbb{T}_{\mathbf{C},f}$ need not be a semi-simple algebra over $\mathbb{C}$.

Lemma 1.34 in the case $A = \mathbb{C}$ says that $V = S_2(\Gamma)^\vee$ is a free $\mathbb{T}_{\mathbf{C}}$-module of rank one, but we also have:

Lemma 1.35 *The module* $S_2(\Gamma)$ *is a free* $\mathbb{T}_{\mathbf{C}}$*-module of rank one.*

Proof: Let $g_1, \ldots, g_t$ be a complete system of newforms of levels $N_1, \ldots, N_t$ dividing N. One can check that the form

$$g = g_1(N/N_1\tau) + \cdots + g_t(N/N_t\tau)$$

generates $S_2(\Gamma)$ as a $\mathbb{T}_{\mathbf{C}}$-module. The map $\mathbb{T}_{\mathbf{C}} \longrightarrow S_2(\Gamma)$ defined by $T \mapsto Tg$ gives an isomorphism from $\mathbb{T}_{\mathbf{C}}$ to $S_2(\Gamma)$, as $\mathbb{T}_{\mathbf{C}}$-modules. □

Remark 1.36 Lemmas 1.34 and 1.35 imply that $\mathbb{T}_{\mathbf{C}}$ is a *Gorenstein* $\mathbb{C}$-algebra (of finite rank), i.e., $\mathrm{Hom}_{\mathbf{C}}(\mathbb{T}_{\mathbf{C}}, \mathbb{C})$ is isomorphic to $\mathbb{T}_{\mathbf{C}}$ as a $\mathbb{T}_{\mathbf{C}}$-module.

Hecke rings over $\mathbb{Q}$: Let $[f]$ be the Galois orbit (under the action of $G_{\mathbb{Q}} = \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$) of a normalized newform f of some level N_f dividing N, and let K_f be the field extension of $\mathbb{Q}$ generated by the Fourier coefficients of f. The space $\oplus_{g \in [f]} S_g$ is a vector space of dimension $[K_f : \mathbb{Q}]\sigma_0(N/N_f)$,

which is spanned by modular forms with rational Fourier coefficients. Let $S_{[f]}$ be the $\mathbb{Q}$-subspace of forms in $\oplus_{g\in[f]}S_g$ with rational Fourier coefficients. The space $S_{[f]}$ is stable under the action of $\mathbb{T}_{\mathbb{Q}}$, and letting $\mathbb{T}_{\mathbb{Q},[f]}$ be the image of $\mathbb{T}_{\mathbb{Q}}$ acting on $S_{[f]}$, we obtain the direct sum decomposition

$$\mathbb{T}_{\mathbb{Q}} = \oplus_{[f]}\mathbb{T}_{\mathbb{Q},[f]},$$

where the sum is taken over the distinct $G_{\mathbb{Q}}$-orbits of normalized newforms f of some level N_f dividing N. If N_f is equal to N, then the algebra $\mathbb{T}_{\mathbb{Q},[f]}$ is isomorphic to the field K_f. If N_f is a proper divisor of N, then, as in the complex case, the algebra $\mathbb{T}_{\mathbb{Q},[f]}$ is a (not necessarily semi-simple) algebra over $\mathbb{Q}$ of rank $\sigma_0(N/N_f)[K_f : \mathbb{Q}]$. The nature of the fields K_f, and in general the structure of $\mathbb{T}_{\mathbb{Q}}$, is very poorly understood at this stage; for example, one does not know how to characterize the number fields that occur as a K_f for some f (but they are all known to be totally real or CM fields).

The ring $\mathbb{T}_{\mathbb{Q}}$ acts naturally on the rational homology $H_1(X_\Gamma, \mathbb{Q}) = \Lambda\otimes\mathbb{Q}$, and we have

Lemma 1.37 *The module $\Lambda \otimes \mathbb{Q}$ is free of rank two over $\mathbb{T}_{\mathbb{Q}}$.*

Sketch of proof: The modules $\Lambda^+ \otimes \mathbb{C} \simeq V$ and $\Lambda^- \otimes \mathbb{C} \simeq V$ are free of rank one over $\mathbb{T}_{\mathbb{C}}$, by lemma 1.34. This implies that $\Lambda^+ \otimes \mathbb{Q}$ and $\Lambda^- \otimes \mathbb{Q}$ are both free of rank one over $\mathbb{T}_{\mathbb{Q}}$. □

Hecke rings over $\mathbb{Z}$: The ring $\mathbb{T}_{\mathbf{Z}}$ is a certain (not necessarily maximal) order in $\mathbb{T}_{\mathbb{Q}}$. One still has an injection

$$\mathbb{T}_{\mathbf{Z}} \hookrightarrow \oplus_{[f]}\mathbb{T}_{\mathbf{Z},[f]},$$

where now $\mathbb{T}_{\mathbf{Z},[f]}$ denotes the ring generated over $\mathbb{Z}$ by the Hecke operators acting on $S_{[f]}$. Of course the structure of $\mathbb{T}_{\mathbf{Z}}$ is even more mysterious than that of $\mathbb{T}_{\mathbb{Q}}$! The ring $\mathbb{T}_{\mathbf{Z}}$ acts naturally on Λ, but it is not the case in general that Λ is free of rank two over $\mathbb{T}_{\mathbf{Z}}$, i.e., that the integral analogue of lemma 1.37 is true. (See remark 1.42 below.)

Hecke rings over $\mathbb{Q}_\ell$: The study of the algebras $\mathbb{T}_{\mathbf{Z}_\ell}$ and $\mathbb{T}_{\mathbb{Q}_\ell}$ arises naturally because of the Hecke action on the Tate module $\mathcal{T}_\ell(J_\Gamma)$,

$$\mathcal{T}_\ell(J_\Gamma) := \varprojlim (J_\Gamma)[\ell^n],$$

where the inverse limit is taken with respect to the multiplication by ℓ maps. The action of $\mathbb{T}_{\mathbf{Z}_\ell}$ on $\mathcal{T}_\ell(J_\Gamma)$ is compatible with that of $G_{\mathbb{Q}}$, and it is this pair of actions on the Tate module which is used to associate two-dimensional Galois representations to modular forms.

It will sometimes be more convenient to consider the ring $\mathbb{T}_{\mathbb{Q}_\ell}$ and its action on

$$\mathcal{V} = \mathcal{T}_\ell(J_\Gamma) \otimes_{\mathbf{Z}_\ell} \mathbb{Q}_\ell.$$

We first record a useful duality property enjoyed by the Tate modules. The Weil pairings on the groups $J_\Gamma[\ell^n]$ for $n \geq 1$ induce a perfect pairing

$$\langle\ ,\ \rangle : T_\ell(J_\Gamma) \times T_\ell(J_\Gamma) \to \mathbb{Z}_\ell.$$

Since each Hecke operator T is adjoint to wTw where $w = w_N$ is the Atkin-Lehner involution, we have the following lemma.

Lemma 1.38 *The map $x \mapsto \phi_x$ where $\phi_x(y) = \langle x, wy\rangle$ defines an isomorphism of $\mathbb{T}_{\mathbb{Z}_\ell}$-modules,*

$$T_\ell(J_\Gamma) \cong T_\ell(J_\Gamma)^\vee = \mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(J_\Gamma), \mathbb{Z}_\ell),$$

and hence an isomorphism of $\mathbb{T}_{\mathbb{Q}_\ell}$-modules

$$\mathcal{V} \cong \mathcal{V}^\vee = \mathrm{Hom}_{\mathbb{Q}_\ell}(\mathcal{V}, \mathbb{Q}_\ell).$$

The following lemma allows us to regard $\mathbb{T}_{\mathbb{Q}_\ell}$ as a coefficient ring for a two-dimensional Galois representation

$$\mathrm{Gal}\,(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}_{\mathbb{T}_{\mathbb{Q}_\ell}}(\mathcal{V}) \cong GL_2(\mathbb{T}_{\mathbb{Q}_\ell}).$$

Lemma 1.39 *The module $\mathcal{V}$ is free of rank 2 over $\mathbb{T}_{\mathbb{Q}_\ell}$.*

Proof: Lemma 1.37 implies directly that the module $\mathcal{V} = \Lambda \otimes \mathbb{Q}_\ell$ is free of rank two over $\mathbb{T}_{\mathbb{Q}_\ell}$. □

Corollary 1.40 *The ring $\mathbb{T}_{\mathbb{Q}_\ell}$ is a Gorenstein $\mathbb{Q}_\ell$-algebra; i.e.,*

$$\mathbb{T}_{\mathbb{Q}_\ell}^\vee = \mathrm{Hom}_{\mathbb{Q}_\ell}(\mathbb{T}_{\mathbb{Q}_\ell}, \mathbb{Q}_\ell)$$

is free of rank one over $\mathbb{T}_{\mathbb{Q}_\ell}$.

Proof: Choosing a basis for $\mathcal{V}$ over $\mathbb{T}_{\mathbb{Q}_\ell}$, we obtain an isomorphism

$$\mathbb{T}_{\mathbb{Q}_\ell} \oplus \mathbb{T}_{\mathbb{Q}_\ell} \cong \mathbb{T}_{\mathbb{Q}_\ell}^\vee \oplus \mathbb{T}_{\mathbb{Q}_\ell}^\vee.$$

Decomposing $\mathbb{T}_{\mathbb{Q}_\ell}$ as $\prod_i R_i$ where each factor R_i is a finite-dimensional local $\mathbb{Q}_\ell$-algebra, we obtain an isomorphism

$$R_i \oplus R_i \cong R_i^\vee \oplus R_i^\vee$$

for each i. At least one of the four maps $R_i^\vee \longrightarrow R_i$ deduced from this isomorphism must be surjective, and by counting dimensions, we see that it must be injective as well. It follows that $\mathbb{T}_{\mathbb{Q}_\ell}$ is isomorphic to $\mathbb{T}_{\mathbb{Q}_\ell}^\vee$. □

Recall that for primes p not dividing N, the Jacobian J_Γ has good reduction mod p, and the Eichler-Shimura relation, theorem 1.29, states that on $J_{\Gamma/\mathbb{F}_p}$, we have

$$T_p = F + \langle p\rangle F'.$$

For primes p not dividing $N\ell$, we may identify $T_\ell(J_\Gamma)$ with the ℓ-adic Tate module of the reduction (see [ST]) and consider the Frobenius endomorphism F on the free rank two $\mathbb{T}_{\mathbb{Q}_\ell}$-module $\mathcal{V}$. As a consequence of the Eichler-Shimura relation, we find:

Theorem 1.41 *For p not dividing $N\ell$, the characteristic polynomial of F on the $\mathbb{T}_{\mathbf{Q}_\ell}$-module $\mathcal{V}$ is*

$$X^2 - T_p X + \langle p \rangle p.$$

Proof: (We are grateful to Brian Conrad for showing us this argument.) Since $FF' = p$, it follows from the Eichler-Shimura relation that

$$F^2 - T_p F + \langle p \rangle p = 0.$$

To conclude that this is in fact the characteristic polynomial, it suffices to compute the trace of F. To do so, we use the $\mathbb{T}_{\mathbf{Q}_\ell}$ isomorphism

$$\mathcal{V} \to \mathcal{V}^\vee$$

defined by the modified pairing $\langle \cdot, w \cdot \rangle$. Under this modified pairing, F is adjoint to $wF'w = \langle p \rangle F'$, so the trace of F on $\mathcal{V}$ is the same as that of $\phi \mapsto \phi \circ (\langle p \rangle F')$ on $\mathcal{V}^\vee$. Choosing bases for $\mathcal{V}$ and $\mathbb{T}^\vee_{\mathbf{Q}_\ell}$, one sees that this is the same as the trace of $\langle p \rangle F'$. Hence

$$2 \operatorname{tr} F = \operatorname{tr} F + \operatorname{tr} (\langle p \rangle F') = \operatorname{tr} (T_p) = 2 T_p.$$

□

Hecke rings over $\mathbb{Z}_\ell$: The ring $\mathbb{T}_{\mathbf{Z}_\ell}$ is free of finite rank over $\mathbb{Z}_\ell$. It therefore decomposes as

$$\mathbb{T}_{\mathbf{Z}_\ell} = \prod \mathbb{T}_{\mathfrak{m}},$$

where the product runs over the maximal ideals $\mathfrak{m}$ of $\mathbb{T}_{\mathbf{Z}_\ell}$, and $\mathbb{T}_{\mathfrak{m}}$ is the localization of $\mathbb{T}_{\mathbf{Z}_\ell}$ at $\mathfrak{m}$. For each $\mathfrak{m}$, $\mathbb{T}_{\mathfrak{m}}$ is a complete local $\mathbb{Z}_\ell$-algebra, free of finite rank as a $\mathbb{Z}_\ell$-module.

Remark 1.42 While the analogue of lemma 1.39 does not always hold for $\mathbb{T}_{\mathbf{Z}_\ell}$ (see [MRi], sec. 13) we shall see that it holds for certain localizations $\mathbb{T}_{\mathfrak{m}}$. Results of this type are much deeper than lemma 1.39 and were first obtained by Mazur [Maz1], sec. 14, 15. We shall return in chapter 4 to explain Mazur's result and its generalizations, which play a role in the arguments of [W3] and [TW].

Example 1.43 The curve $X_0(19)$ has genus 1, and $X_0(57)$ has genus 5. By consulting the tables in the Antwerp volume [Ant4] or Cremona's book [Cr], one finds that there is exactly one newform of level 19, and that there are three newforms of level 57, which all have rational Fourier coefficients. Their Fourier coefficients a_p, for the first few primes p, are listed in the following table:

	2	3	5	7	11	13	17	19	23	29	31
19*A*	0	−2	3	−1	3	−4	−3	1	0	6	−4
57*A*	−2	−1	−3	−5	1	2	−1	−1	−4	−2	−6
57*B*	1	1	−2	0	0	6	−6	−1	4	2	8
57*C*	−2	1	1	3	−3	−6	3	−1	4	−10	2

Setting $f = 19A$, we find that a basis of simultaneous eigenforms in $S_2(57, \mathbb{C})$ for the Hecke operators T_p, $l \neq 3, 19$, and U_3, U_{19} is:

$$f(\tau) + (1+\sqrt{-2})f(3\tau), \quad f(\tau) + (1-\sqrt{-2})f(3\tau), \quad 57A, \quad 57B, \quad 57C.$$

It appears from the table that the Fourier coefficients corresponding to the forms $57B$ and $57C$ are congruent modulo 3. This is in fact the case. One finds that the Hecke ring $\mathbb{T}_{\mathbb{Z}_3}$ generated over $\mathbb{Z}_3$ by the Hecke operators acting on $S_2(57, \mathbb{Z}_3)$ is isomorphic to the subalgebra of $\mathbb{Z}_3^5$:

$$\{(x, y, z, t, w) \text{ such that } t \equiv w \pmod 3\}.$$

The isomorphism sends T_p (for $p \neq 3, 19$) to the element

$$(a_p(19A), a_p(19A), a_p(57A), a_p(57B), a_p(57C)).$$

It sends U_3 to $(-1+\sqrt{-2}, -1-\sqrt{-2}, -1, 1, 1)$ and U_{19} to $(1, 1, -1, -1, -1)$. Thus there are exactly 4 distinct maximal ideals of $\mathbb{T}_{\mathbb{Z}_3}$, and the localizations at these maximal ideals are the rings $\mathbb{Z}_3$, $\mathbb{Z}_3$, $\mathbb{Z}_3$, and

$$\mathbb{T}_{\mathfrak{m}} = \{(t, w) \text{ such that } t \equiv w \pmod 3\},$$

where $\mathfrak{m}$ is the ideal generated by 3 and $T_n - a_n(57B)$ for all n.

1.7 The Shimura construction

Let $f = \sum a_n q^n$ be an eigenform on Γ with (not necessarily rational) Fourier coefficients, corresponding to a surjective algebra homomorphism

$$\lambda_f : \mathbb{T}_{\mathbb{Q}} \longrightarrow K_f,$$

where K_f is the field generated over $\mathbb{Q}$ by the Fourier coefficients of f. We briefly review in this section a construction of Shimura ([Shi2], ch. 7) which associates to f (or rather, to the orbit $[f]$ of f under $G_{\mathbb{Q}}$) an abelian variety A_f defined over $\mathbb{Q}$ and of dimension $[K_f : \mathbb{Q}]$.

Let $I_f \subset \mathbb{T}_{\mathbb{Z}}$ be the ideal $\ker(\lambda_f) \cap \mathbb{T}_{\mathbb{Z}}$. The image $I_f(J_\Gamma)$ is a (connected) subabelian variety of J_Γ which is stable under $\mathbb{T}_{\mathbb{Z}}$ and is defined over $\mathbb{Q}$.

Definition 1.44 The abelian variety A_f associated to f is the quotient

$$A_f = J_\Gamma / I_f J_\Gamma$$

From this definition one sees that A_f is defined over $\mathbb{Q}$ and depends only on $[f]$, and that its endomorphism ring contains $\mathbb{T}_{\mathbb{Z}}/I_f$ which is isomorphic to an order in K_f.

Remark 1.45 Using theorem 1.22, one can show that $J_0(N)$ is isogenous to $\prod_{[f]} A_{[f]}^{\sigma_0(N/N_f)}$.

We now describe the abelian variety A_f as a complex torus. Let V_f be the subspace of $V = S_2(\Gamma)^\vee$ on which $\mathbb{T}$ acts by λ_f. Theorem 1.22 and lemma 1.34 show that V_f is a one-dimensional complex vector space. Let π_f be the orthogonal projection of V to V_f, relative to the Petersson scalar product. The projector π_f belongs naturally to $\mathbb{T}_{K_f}$.

Let $[f]$ be the set of all eigenforms whose Fourier coefficients are Galois conjugate to those of f. The number of forms in $[f]$ is equal to the degree d of K_f over $\mathbb{Q}$. Now, we set

$$V_{[f]} = \oplus_{g \in [f]} V_g, \quad \pi_{[f]} = \sum_{g \in [f]} \pi_g.$$

Note that $\pi_{[f]}$ is simply the orthogonal projection of V to $V_{[f]}$. Note also that $\pi_{[f]}$ belongs to the Hecke algebra $\mathbb{T}_{\mathbb{Q}}$.

Lemma 1.46 *The abelian variety A_f is isomorphic over $\mathbb{C}$ to the complex torus $V_{[f]}/\pi_{[f]}(\Lambda)$, with the map $\pi_{[f]} : V/\Lambda \longrightarrow V_{[f]}/\pi_{[f]}(\Lambda)$ corresponding to the natural projection from J_Γ to A_f.*

In particular, one sees that A_f is an abelian variety of dimension $d = [K_f : \mathbb{Q}]$. Hence if f has rational Fourier coefficients, then the abelian variety A_f is an elliptic curve. This elliptic curve is called the *strong modular elliptic curve* associated to f if also f is a newform of level N and $\Gamma = \Gamma_0(N)$.

Example 1.47 If $\Gamma = \Gamma_0(26)$, one checks that the genus of X_Γ is two, and that there are two distinct normalized eigenforms f_1 and f_2 in $S_2(26)$. From the tables in [Ant4] or [Cr], one sees that f_1 and f_2 have integral Fourier coefficients, whose values for the primes ≤ 31 are:

	2	3	5	7	11	13	17	19	23	29	31
f_1	−1	1	−3	−1	6	1	−3	2	0	6	−4
f_2	1	−3	−1	1	−2	−1	−3	6	−4	2	4

Hence the abelian varieties A_{f_1} and A_{f_2} are elliptic curves. The above table suggests (and this can be checked directly by looking at the equations for these curves given in [Ant4] or [Cr], or by using the discussion in [DO], lemma 2.1) that the Fourier coefficients of f_1 and f_2 are congruent modulo 2. The natural projection $J_0(26) \longrightarrow A_{f_1} \oplus A_{f_2}$ is an isogeny whose kernel is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, and $J_0(26)$ is *not* isomorphic to $A_{f_1} \oplus A_{f_2}$. More generally, one knows that a Jacobian can never decompose as a non-trivial direct product of two principally polarized abelian varieties, cf. [Maz1], prop. 10.6. (There are no non-zero homomorphisms from A_{f_1} to A_{f_2}, and so a non-trivial product decomposition of J_Γ would have to induce a decomposition as a product of principally polarized abelian varieties.)

Let $T_\ell(A_f)$ be the Tate module of the abelian variety A_f,

$$T_\ell(A_f) := \varprojlim (A_f)[\ell^n],$$

where the inverse limit is taken with respect to the multiplication by ℓ maps. This module is naturally a module for $\mathbb{T}_f \otimes \mathbb{Z}_\ell$, and $T_\ell(A_f) \otimes \mathbb{Q}_\ell$ is a module for $K_f \otimes \mathbb{Q}_\ell$ as well.

Lemma 1.48 *The module $T_\ell(A_f) \otimes \mathbb{Q}_\ell$ is a free module of rank 2 over $K_f \otimes \mathbb{Q}_\ell$.*

Proof: This follows directly from lemma 1.39. □

Proposition 1.49 *The algebra $\mathrm{End}_{\mathbb{Q}}(A_f) \otimes \mathbb{Q}$ is isomorphic to K_f. In particular A_f is simple over $\mathbb{Q}$.*

The proof is given in [R2], cor. 4.2. The main ingredient is the *irreducibility* of the Galois representation attached to A_f as in section 3.1. (Cf. theorem 3.1.)

Properties of A_f: good reduction:

Theorem 1.50 *If f is an eigenform of level N, and p is a prime that does not divide N, then the abelian variety A_f has good reduction at p.*

Proof: This follows from the fact that J_Γ has good reduction at such primes, which in turn is a consequence of the good reduction of X_Γ. □

If p is a prime not dividing N, it then becomes natural to study the number $N_{f,p}$ of points on the abelian variety A_f over the finite field $\mathbb{F}_p$. It is given by the following formula:

Proposition 1.51 *The number of points $N_{f,p}$ is given by the formula*

$$N_{f,p} = \mathrm{Norm}_{K_f/\mathbb{Q}}(\lambda_f(1 - a_p(f) + \langle p \rangle p)),$$

where $a_f(p) \in K_f$ is the p-th Fourier coefficient of f.

Proof: By Weil's theory [Wel], this number is given by the determinant

$$N_{f,p} = \det(1 - F),$$

where F is the Frobenius endomorphism acting on the ℓ-adic Tate module $T_\ell(A_f)$. So theorem 1.41 implies

$$\det(1 - F) = \mathrm{Norm}_{K_f/\mathbb{Q}}(\lambda_f(1 - T_p + \langle p \rangle p)),$$

and proposition 1.51 follows. □

Defining the local Hasse-Weil L-function of A_f over $\mathbb{F}_p$ by the formula

$$L(A_f/\mathbb{F}_p, s) = \det(1 - Fp^{-s})^{-1},$$

the proof above gives the formula:

$$L(A_f/\mathbb{F}_p, s) = \prod_\sigma L_p(f^\sigma, s),$$

where the product is taken over all complex embeddings $\sigma : K_f \hookrightarrow \mathbb{C}$, and $L_p(f^\sigma, s)$ is the Euler factor at p in the L-function that was associated to f^σ in section 1.4.

In particular, if $A_f = E$ is an elliptic curve, i.e., f has rational Fourier coefficients and f is on $\Gamma_0(N)$, then the number of points on E over $\mathbb{F}_p$ is given by the formula

$$\#E(\mathbb{F}_p) = p + 1 - a_p(f). \tag{1.7.1}$$

Properties of A_f: bad reduction: A fundamental quantity associated to any abelian variety A over $\mathbb{Q}$ is its *arithmetic conductor*, which measures the amount of bad reduction of A. If we factor this conductor as a product of prime powers, $\prod_p p^{m_p}$, then the exponents m_p are equal to 0 precisely when A has good reduction at p. The definition of $m_p(A)$ was already given in section 1.1 when A is an abelian variety of dimension 1, i.e., an elliptic curve. In general, the exponent m_p coincides with $m_p(\rho_{A,\ell})$ (see section 2.1 below), where $\rho_{A,\ell}$ is the Galois representation on the ℓ-adic Tate module of A.

Regarding the bad reduction of the abelian variety A_f, one has the following consequence of the results of Langlands, Deligne and Carayol [Ca3] discussed below in section 3.1.

Theorem 1.52 *If f is a newform of level N, then the conductor of the abelian variety A_f is equal to N^g.*

1.8 The Shimura-Taniyama conjecture

Let E be any elliptic curve defined over $\mathbb{Q}$, and let N denote its arithmetic conductor, defined as in section 1.1. Then we have:

Proposition 1.53 *The following are equivalent:*

(a) *The curve E is isogenous over $\mathbb{Q}$ to A_f, for some newform f on some congruence group Γ.*

(b) *The curve E is isogenous over $\mathbb{Q}$ to A_f, for a newform f on $\Gamma_0(N)$.*

(c) *There is a non-constant morphism defined over $\mathbb{Q}$, from $X_0(N)$ to E.*

Sketch of proof: The implication $(b) \Rightarrow (a)$ is immediate, and $(a) \Rightarrow (b)$ follows from the work of Carayol [Ca3] (cf. proposition 3.20 below). If (b) holds, then there is a surjective map $J_0(N) \longrightarrow E$. Composing with the Abel-Jacobi map $X_0(N) \longrightarrow J_0(N)$ gives a map to E which is not constant, by the definition of the Jacobian, and hence (c) holds. Conversely, a non-constant map of curves $\pi : X_0(N) \longrightarrow E$ induces, by Albanese (covariant) functoriality, a surjective map of Jacobians $\pi_* : J_0(N) \longrightarrow E$, where we have identified E with its own Jacobian in the natural way. Since $J_0(N)$

is isogenous to a product (possibly with multiplicities) of abelian varieties A_f where f run over newforms of level N_f dividing N (cf. remark 1.45), it follows that there is a surjective map $A_f \longrightarrow E$ for some A_f. Since A_f is simple (proposition 1.49), this map is an isogeny, and part (b) follows. □

We call an elliptic curve over $\mathbb{Q}$ satisfying the equivalent properties above a *modular elliptic curve.* A startling conjecture that was first proposed by Taniyama in the 1950's and made more precise by Shimura predicts that the Shimura construction explained in the previous section is surjective:

Conjecture 1.54 *All elliptic curves defined over $\mathbb{Q}$ are modular.*

This conjecture is now known to be true for a very wide class of elliptic curves. The results of Wiles [W3] and Taylor-Wiles [TW] imply it is true for all semi-stable elliptic curves, and a strengthening of the method, [Di2], implies the conjecture for all elliptic curves which have semi-stable reduction at 3 and 5.

Remark 1.55 See [DI], sec. 13 for a more thorough list of equivalent forms of the conjecture.

Relation with L-functions: Define numbers $a_p = a_p(E)$ as in section 1.1. Recall the (global) Hasse-Weil L-function defined in section 1.1 by equations (1.1.2), (1.1.4), (1.1.5), and (1.1.6). The following proposition reveals some of the importance of the Shimura-Taniyama conjecture:

Proposition 1.56 *If E is modular, i.e., is associated to a newform f by the Shimura construction, then*

$$L(E/\mathbb{Q}, s) = L(f, s).$$

In particular, $L(E/\mathbb{Q}, s)$ has an analytic continuation to the entire complex plane, and a functional equation.

Sketch of proof: If E is isogenous to an elliptic curve A_f associated to a newform f on $\Gamma_0(N)$ by the Shimura construction, then the two L-functions $L(A_f/\mathbb{Q}, s)$ and $L(E/\mathbb{Q}, s)$ are equal. On the other hand, formula (1.7.1) directly implies that $L(A_f/\mathbb{Q}, s)$ is equal to $L(f, s)$, at least up to finitely many Euler factors (corresponding precisely, by the work of Igusa, to the primes dividing N). The equality of all the Euler factors follows from the work of Deligne, Langlands, and Carayol (cf. [Ca1]). □

Knowing the analytic continuation and functional equation of the Hasse-Weil L-function of an elliptic curve is of great importance in the theory. For example, the conjecture of Birch and Swinnerton-Dyer (which was stated in a weak form in section 1.1, conjecture 1.9) relates arithmetic invariants of E such as the rank of its Mordell-Weil group and the order of its Shafarevich-Tate group to the behaviour at $s = 1$ of $L(E/\mathbb{Q}, s)$. (Note that $s = 1$ is

outside the domain of absolute convergence of the infinite product used to define $L(E/\mathbb{Q}, s)$.)

One consequence of this conjecture is that $E(\mathbb{Q})$ is finite if $L(E/\mathbb{Q}, 1)$ is non-zero. This was proved by Coates and Wiles in [CW] for elliptic curves with complex multiplication, and by Gross-Zagier and Kolyvagin [Kol] for modular elliptic curves. (Recently, a different proof for modular elliptic curves has been announced by K. Kato.) Thanks to the breakthroughs of [W3], [TW], and [Di2], the results of by Gross-Zagier, Kolyvagin and Kato are now unconditional for a very large class of elliptic curves.

The Shimura-Taniyama conjecture for abelian varieties: We may view conjecture 1.54 as asserting that the map

$$\left\{\begin{array}{l}\text{Newforms of weight 2 on } X_0(N)\\ \text{with rational Fourier coefficients}\end{array}\right\} \longrightarrow \left\{\begin{array}{c}\text{Isogeny classes of elliptic curves}\\ \text{over } \mathbb{Q} \text{ of conductor N}\end{array}\right\}$$

provided by the Shimura construction is a bijection. It is natural to extend this result to all eigenforms, not just those having rational Fourier expansions. We say that an abelian variety defined over $\mathbb{Q}$ is *modular* if it is isogenous to an abelian variety of the form A_f for some newform f on $\Gamma_1(N)$. It would not be reasonable to expect that all abelian varieties over $\mathbb{Q}$ are modular: the abelian varieties arising from the Shimura construction are very special, in many respects. Most importantly, they have a very large ring of endomorphisms. Following Ribet [R7], we make the definition:

Definition 1.57 An abelian variety over $\mathbb{Q}$ of dimension g is said to be of GL_2-type if its endomorphism ring over $\mathbb{Q}$ contains an order in a field of degree g over $\mathbb{Q}$.

Then we have the following generalized Shimura-Taniyama conjecture:

Conjecture 1.58 *Every simple abelian variety A over $\mathbb{Q}$ which is of GL_2-type is modular.*

This generalization of the Shimura-Taniyama conjecture is far from being proved, and tackling it still seems to require some major new ideas, even after the ideas introduced by Wiles.

2 Galois theory

2.1 Galois representations

If F is a perfect field, we will let $\bar{F}$ denote an algebraic closure of F and $G_F = \mathrm{Gal}(\bar{F}/F)$ its absolute Galois group. Recall that G_F is a profinite group: more precisely, $G_F = \varprojlim \mathrm{Gal}(L/F)$ as L runs over finite Galois extensions of F contained in $\bar{F}$. Hence G_F carries a natural topology. If ℓ is a prime different from the characteristic of F we will let $\epsilon_\ell : G_F \to \mathbb{Z}_\ell^\times$

denote the ℓ-adic cyclotomic character, i.e. if ζ is an ℓ-power root of unity in $\bar{F}$ then $\sigma(\zeta) = \zeta^{\epsilon_\ell(\sigma)}$ for all $\sigma \in G_F$. We will simply write ϵ if the choice of ℓ is clear from the context.

The finite fields $\mathbb{F}_p$: The group $G_{\mathbb{F}_p}$ is isomorphic to $\hat{\mathbb{Z}}$, the profinite completion of $\mathbb{Z}$. A natural topological generator is provided by the Frobenius element Frob_p, the automorphism of $\bar{\mathbb{F}}_p$ which raises elements of $\bar{\mathbb{F}}_p$ to their p^{th} power. If ℓ is different from p, then $\epsilon_\ell(\mathrm{Frob}_p) = p$.

The p-adic fields $\mathbb{Q}_p$: The p-adic valuation $v_p : \mathbb{Q}_p^\times \twoheadrightarrow \mathbb{Z}$ extends uniquely to a valuation $v_p : \bar{\mathbb{Q}}_p^\times \twoheadrightarrow \mathbb{Q}$. Let $\mathcal{O}_{\bar{\mathbb{Q}}_p}$ denote the ring of integers of $\bar{\mathbb{Q}}_p$ and $\mathfrak{m}_{\bar{\mathbb{Q}}_p}$ its maximal ideal. The residue field $\mathcal{O}_{\bar{\mathbb{Q}}_p}/\mathfrak{m}_{\bar{\mathbb{Q}}_p}$ is an algebraic closure of $\mathbb{F}_p$, which we will identify with $\bar{\mathbb{F}}_p$. The valuation v_p is compatible with the action of $G_{\mathbb{Q}_p}$, and hence $\mathcal{O}_{\bar{\mathbb{Q}}_p}$ and $\mathfrak{m}_{\bar{\mathbb{Q}}_p}$ are stable under $G_{\mathbb{Q}_p}$. In particular $G_{\mathbb{Q}_p}$ acts on $\mathcal{O}_{\bar{\mathbb{Q}}_p}/\mathfrak{m}_{\bar{\mathbb{Q}}_p}$ and we obtain a map $\varrho : G_{\mathbb{Q}_p} \to G_{\mathbb{F}_p}$, which is in fact a surjection. We call the kernel the *inertia group* and denote it I_p. It has a unique maximal pro-p subgroup: the *wild inertia group*, denoted P_p. There is a canonical isomorphism

$$I_p/P_p \cong \prod_{\ell \neq p} \mathbb{Z}_\ell(1),$$

where the (1) indicates that if $f \in G_{\mathbb{Q}_p}/P_p$ is a lifting of Frobenius and if $\sigma \in I_p/P_p$ then $f\sigma f^{-1} = \sigma^p$.

The inertia group I_p is filtered by a series of subgroups called the higher ramification groups. More precisely, there are closed normal subgroups I_p^u in $G_{\mathbb{Q}_p}$ for any $u \in [-1, \infty]$, with the following properties:

- If $u \leq v$ then $I_p^u \supset I_p^v$, i.e., the I_p^u form a *decreasing filtration.*
- For $u \leq 0$, $I_p^u = I_p$ while $I_p^\infty = \{1\}$.
- $P_p = \bigcup_{u>0} I_p^u$.
- $I_p^u = \bigcap_{v<u} I_p^v$.

These groups are defined in section 3 of chapter IV of [Se2], where they are denoted $G_{\mathbb{Q}_p}^u$ (except that there, $G_{\mathbb{Q}_p}^{-1} = G_{\mathbb{Q}_p}$ whereas we have $I_p^{-1} = I_p$).

The local Kronecker-Weber theorem ([Se2], ch. 14, sec. 7, for example) asserts that the map

$$\varrho \times \epsilon_p : G_{\mathbb{Q}_p}^{\mathrm{ab}} \longrightarrow G_{\mathbb{F}_p} \times \mathbb{Z}_p^\times.$$

is an isomorphism. (We will use G^{ab} to denote the abelianisation of a profinite group G, i.e. the unique maximal abelian continuous image.) Under it I_p goes to $\mathbb{Z}_p^\times$ and for $u > 0$, I_p^u goes to $(1 + p^{\lceil u \rceil}\mathbb{Z}_p) \subset \mathbb{Z}_p^\times$, where $\lceil u \rceil$ denotes the least integer greater than or equal to u. If $\ell \neq p$ then ϵ_ℓ is trivial on I_p and takes Frob_p to $p \in \mathbb{Z}_\ell^\times$.

The field $\mathbb{Q}$: If p is a rational prime, the usual p-adic valuation $v_p : \mathbb{Q}^\times \to \mathbb{Z}$ extends to a valuation $v : \bar{\mathbb{Q}}^\times \to \mathbb{Q}$. This extension is not unique, but all extensions are permuted transitively by $G_{\mathbb{Q}}$. Fix one such extension v_p and let G_p denote its stabiliser in $G_{\mathbb{Q}}$. Then G_p acts on the completion $(\bar{\mathbb{Q}})_{v_p}$ and preserves the subfield of elements algebraic over $\mathbb{Q}_p$. This field is an algebraic closure of $\mathbb{Q}_p$, which we shall identify with $\bar{\mathbb{Q}}_p$. One can check that the resulting map $G_p \to G_{\mathbb{Q}_p}$ is an isomorphism. Thus we obtain subgroups $G_p \supset I_p \supset I_p^u$ for $u \in [-1, \infty]$ and a distinguished element $\mathrm{Frob}_p \in G_p/I_p$. These objects depend on the choice of v_p and vary by conjugation in $G_{\mathbb{Q}}$ as this choice is varied. We call an algebraic extension $F/\mathbb{Q}$ *unramified at p* if all conjugates of I_p lie in $G_F \subset G_{\mathbb{Q}}$; otherwise we say that $F/\mathbb{Q}$ is *ramified.* If $F/\mathbb{Q}$ is Galois and unramified at p, then there is a well-defined conjugacy class $[\mathrm{Frob}_p] \subset \mathrm{Gal}(F/\mathbb{Q})$.

By replacing the p-adic completion by an Archimedean completion, one has a well-defined conjugacy class $[c]$ in $G_{\mathbb{Q}}$ consisting of those elements that arise as complex conjugation for some embedding $\bar{\mathbb{Q}} \hookrightarrow \mathbb{C}$. We will denote by G_∞ the subgroup $\{1, c\}$ for one such element c.

We have the following fundamental results concerning the structure of $G_{\mathbb{Q}}$ (see for instance [La1]).

Theorem 2.1 *If $F/\mathbb{Q}$ is a finite extension then F is only ramified at finitely many primes (those dividing the discriminant of $F/\mathbb{Q}$).*

Theorem 2.2 (Hermite-Minkowski) *If S is a finite set of primes and if $d \in \mathbb{Z}_{>0}$ then there are only finitely many extensions $F/\mathbb{Q}$ of degree d which are unramified outside S (contained in a fixed algebraic closure $\bar{\mathbb{Q}}$).*

Theorem 2.3 (Chebotarev) *If $F/\mathbb{Q}$ is a Galois extension unramified outside a finite set S of primes then $\bigcup_{p \notin S}[\mathrm{Frob}_p]$ is dense in $\mathrm{Gal}(F/\mathbb{Q})$.*

Theorem 2.4 (Kronecker-Weber) *The product of the p-adic cyclotomic characters gives an isomorphism*

$$\prod_p \epsilon_p : G_{\mathbb{Q}}^{\mathrm{ab}} \xrightarrow{\sim} \prod_p \mathbb{Z}_p^\times.$$

We remark that the Chebotarev density theorem (theorem 2.3) applied to the extension of $\mathbb{Q}$ generated by the m^{th} roots of unity implies Dirichlet's theorem that if n and m are coprime integers then there are infinitely many primes p with $p \equiv n \bmod m$. We also remark that the Kronecker-Weber theorem is equivalent to the fact that the maximal abelian extension of $\mathbb{Q}$ is obtained by adjoining all roots of unity to $\mathbb{Q}$.

Representations: A d-dimensional representation of $G_{\mathbb{Q}}$ is a homomorphism

$$G_{\mathbb{Q}} \longrightarrow GL_d(K),$$

where K is any field. (In our later discussion, we will also consider representations with coefficients in a ring.) Often K (and hence $GL_d(K)$) comes equipped with a natural topology, and it is then customary to restrict one's attention to continuous homomorphisms $G_{\mathbb{Q}} \longrightarrow GL_d(K)$.

Since any one-dimensional representation has abelian image, theorem 2.4 allows one to give a complete description of the one-dimensional representations of $G_{\mathbb{Q}}$ together with the behaviour of these representations on the decomposition groups at all primes. The aim of this article is to discuss attempts to give a similar theory for the two-dimensional representations of $G_{\mathbb{Q}}$.

We will call a representation ρ of $G_{\mathbb{Q}}$ unramified at p if it is trivial on the inertia group I_p. Otherwise we say that it is ramified at p. If ρ is unramified at p, then $\rho(\mathrm{Frob}_p)$ is well-defined (and its conjugacy class is independent of the choice of v_p).

If ρ is a representation of a group G and i is a non-negative integer, then we let $\wedge^i\rho$ denote the representation of G on the i^{th} exterior power of the underlying module of ρ. If H is a subgroup of G, we will let ρ^H (resp. ρ_H) denote the H-invariants (resp. H-coinvariants) of the underlying module of ρ. If H is normal in G, then we shall also use ρ^H and ρ_H to denote the corresponding representation of G/H. In particular if p is a prime and ρ is a representation of $G_{\mathbb{Q}}$ or $G_{\mathbb{Q}_p}$, then $\rho^{I_p}(\mathrm{Frob}_p)$ and $\rho_{I_p}(\mathrm{Frob}_p)$ are well-defined.

We will be primarily interested in three types of representation of $G_{\mathbb{Q}}$.

- **Artin representations**, i.e. continuous representations $G_{\mathbb{Q}} \to GL_d(\mathbb{C})$. Since all compact totally disconnected subgroups of $GL_d(\mathbb{C})$ are finite, Artin representations have finite image. Hence they are semi-simple and are unramified at all but finitely many primes, by theorem 2.1.

- **Mod ℓ representations**, i.e. continuous representations $G_{\mathbb{Q}} \to GL_d(k)$, where k is a finite field of characteristic ℓ. These always have finite image and hence, like Artin representations, are unramified at all but finitely many primes.

- **ℓ-adic representations**, i.e. continuous representations $G_{\mathbb{Q}} \to GL_d(K)$, where K is a finite extension of $\mathbb{Q}_\ell$. We require that an ℓ-adic representation be unramified at all but finitely many primes.

Remark 2.5 (a) Continuous representations $G_{\mathbb{Q}} \to GL_d(K)$, unlike those to $GL_d(\mathbb{C})$ or $GL_d(k)$, may be ramified at infinitely many primes. We shall not need to consider such representations. (One rarely if ever does.) However, the term "ℓ-adic representation" is often used in the more general sense elsewhere in the literature.

(b) Note that the image of an ℓ-adic representation can very well be infinite. For instance this is the case if $d = 1$ and the representation is the cyclotomic character.

(c) Note that in the case of mod ℓ and ℓ-adic representations, the representations need not be semi-simple.

Proposition 2.6 *Let S be any finite set of primes.*

(a) *An Artin representation $\rho : G_{\mathbb{Q}} \to GL_d(\mathbb{C})$ is determined by the values of* $\operatorname{tr} \rho(\mathrm{Frob}_p)$ *on the primes $p \notin S$ at which ρ is unramified.*

(b) *A semi-simple mod ℓ representation $\rho : G_{\mathbb{Q}} \to GL_d(k)$ is determined by the values of* $\operatorname{tr} \wedge^i \rho(\mathrm{Frob}_p)$ *$(i = 1, \ldots, d)$ on the primes $p \notin S$ at which ρ is unramified. If $\ell > d$ then it is determined by* $\operatorname{tr} \rho(\mathrm{Frob}_p)$ *on the primes $p \notin S$ at which ρ is unramified.*

(c) *A semi-simple ℓ-adic representation $\rho : G_{\mathbb{Q}} \to GL_d(K)$ is determined by the values of* $\operatorname{tr} \rho(\mathrm{Frob}_p)$ *on the primes $p \notin S$ at which ρ is unramified.*

Proof: Combining the Chebotarev density theorem (theorem 2.3) with the continuity of ρ we see that we must show that $\operatorname{tr} \rho$ (resp. $\operatorname{tr} \wedge^i \rho$) determines ρ up to conjugacy in the various settings. For the characteristic zero representations, see [Bour], ch. 8, sec. 12.1, prop. 3; for mod ℓ representations, this is the Brauer-Nesbitt theorem ([CR], (30.16)). □

Let K denote a finite extension of $\mathbb{Q}_\ell$, let $\mathcal{O}$ be its ring of integers, λ the maximal ideal of $\mathcal{O}$ and k the residue field. If $\rho : G_{\mathbb{Q}} \to GL_d(K)$ is an ℓ-adic representation, then the image of ρ is compact, and hence ρ can be conjugated to a homomorphism $G_{\mathbb{Q}} \to GL_d(\mathcal{O})$. Reducing modulo the maximal ideal λ gives a *residual representation* $\bar{\rho} : G_{\mathbb{Q}} \to GL_d(k)$. This representation may depend on the particular $GL_d(K)$-conjugate of ρ chosen, but its semi-simplification $\bar{\rho}^{ss}$ (i.e. the unique semi-simple representation with the same Jordan-Hölder factors) is uniquely determined by ρ, by proposition 2.6 (b).

Note that the kernel of the reduction map $GL_d(\mathcal{O}) \to GL_d(k)$ is a pro-ℓ-group. In particular we see that if $p \neq \ell$ then $\rho(P_p)$ is finite and $\rho(P_p)$ is isomorphic to $\bar{\rho}(P_p)$.

Suppose ρ is a representation of $G_{\mathbb{Q}}$ of one of the three sorts above. If $p \neq \ell$ then we define the conductor, $m_p(\rho)$, of ρ at p by

$$m_p(\rho) = \int_{-1}^{\infty} \operatorname{codim} \rho^{I_p^u} \, du = \operatorname{codim} \rho^{I_p} + \int_{0}^{\infty} \operatorname{codim} \rho^{I_p^u} \, du.$$

This is well-defined as $\rho(P_p)$ is a finite group. If ρ is an Artin representation this makes sense also for $p = \ell$. (It would even make sense for a mod ℓ representation if $p = \ell$, but in this case it does not seem to be a useful notion.) It is known that $m_p(\rho)$ is an integer (see chapter VI of [Se2]). Moreover it is easily seen that $m_p(\rho) = 0$ if and only if ρ is unramified at p. We define the conductor $N(\rho)$ of ρ to be

$$\prod_p p^{m_p(\rho)},$$

where the product is over all primes $p \neq \ell$ in the case of an ℓ-adic or mod ℓ representation and over all primes in the case of an Artin representation. This makes sense because ρ is unramified almost everywhere.

The following lemma is an exercise.

Lemma 2.7 *Suppose that $\rho : G_{\mathbb{Q}} \to GL_d(K)$ is an ℓ-adic representation and that $\bar{\rho} : G_{\mathbb{Q}} \to GL_d(k)$ is a reduction of ρ. Then for each prime $p \neq \ell$ we have $\int_0^\infty \mathrm{codim} \rho^{I_p^u}\, du = \int_0^\infty \mathrm{codim} \bar{\rho}^{I_p^u}\, du$. Thus*

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

and in particular $N(\bar{\rho})$ divides $N(\rho)$.

We take this opportunity to introduce some important notation. Let G denote a group and R a ring. If $\rho : G \to GL_d(R)$ is a representation we shall let M_ρ denote the underlying $R[G]$-module (so that $M_\rho \cong R^d$ as an R-module). If M is an $R[G]$-module we shall let $\mathrm{End}\,(M)$ denote the module of R-linear endomorphisms of M. It is also an $R[G]$-module with the G-action being defined by

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

If M is a finitely generated free R-module we will let $\mathrm{End}^0(M)$ denote the sub-$R[G]$-module of $\mathrm{End}\,(M)$ consisting of endomorphisms of trace zero. We will use $\mathrm{ad}\rho$ to denote $\mathrm{End}\,(M_\rho)$ and $\mathrm{ad}^0\rho$ to denote $\mathrm{End}^0(M_\rho)$. Note that if d is invertible in R then $\mathrm{ad}\rho \cong \mathrm{ad}^0\rho \oplus R$ as $R[G]$-modules, where R has the trivial action of G. Note also then that the kernel of $\mathrm{ad}^0\rho$ is the same as the kernel of the composite map $G \xrightarrow{\rho} GL_d(R) \to PGL_d(R)$.

We also remark that if R is an algebraically closed field of characteristic other than 2 and $\rho : G \to GL_2(R)$ is irreducible, then either $\mathrm{ad}^0\rho$ is irreducible or there is a subgroup $H \subset G$ of index 2 and a character $\chi : H \to R^\times$ such that $\rho \cong \mathrm{Ind}_H^G \chi$. In the latter case $\mathrm{ad}^0\rho \cong \delta \oplus \mathrm{Ind}_H^G(\chi/\chi')$, where δ is the nontrivial character of G/H and χ' is the composite of χ and conjugation by an element of $G - H$. Moreover, either $\mathrm{Ind}_H^G(\chi/\chi')$ is irreducible or $\mathrm{ad}^0\rho \cong \delta_1 \oplus \delta_2 \oplus \delta_3$ where the δ_i are distinct characters of G of order 2. In the latter case, ρ is induced from a character from each of the subgroups $H_i = \ker \delta_i$ of index 2 in G. We leave the verification of these facts as an exercise to the reader.

2.2 Representations associated to elliptic curves

Perhaps the simplest examples of non-abelian ℓ-adic representations arise from elliptic curves defined over $\mathbb{Q}$.

We will let $E[n](\bar{\mathbb{Q}})$ denote the group of n-torsion points on $E(\bar{\mathbb{Q}})$. By proposition 1.1 there is a non-canonical isomorphism $E[n](\bar{\mathbb{Q}}) \cong (\mathbb{Z}/n\mathbb{Z})^2$.

Furthermore, $E[n](\bar{\mathbb{Q}})$ carries a natural action of $G_{\mathbb{Q}}$ and so we get a representation (defined up to conjugation)

$$\bar{\rho}_{E,n} : G_{\mathbb{Q}} \to GL_2(\mathbb{Z}/n\mathbb{Z}).$$

If ℓ is a prime then we set $T_\ell E = \varprojlim E[\ell^n](\bar{\mathbb{Q}})$, which is non-canonically isomorphic to $\mathbb{Z}_\ell^2$. Again it has a natural continuous action of $G_{\mathbb{Q}}$, and so we get a representation (defined up to conjugation)

$$\rho_{E,\ell} : G_{\mathbb{Q}} \to GL_2(\mathbb{Z}_\ell).$$

Note that $\bar{\rho}_{E,\ell^n} \cong (\rho_{E,\ell} \bmod \ell^n)$.

Global properties: We have the following basic properties of the representations $\bar{\rho}_{E,\ell^n}$ and $\rho_{E,\ell}$:

Proposition 2.8 (a) *The determinant of $\rho_{E,\ell}$ is ϵ_ℓ.*

(b) *The representation $\rho_{E,\ell}$ is absolutely irreducible for all ℓ. For fixed E, $\bar{\rho}_{E,\ell}$ is absolutely irreducible for all but finitely many ℓ.*

(c) *If E does not have complex multiplication then $\rho_{E,\ell}$ (and hence $\bar{\rho}_{E,\ell}$) is surjective for all but finitely many ℓ.*

Proof: Part (a) follows from the existence of the non-degenerate alternating Galois-equivariant Weil pairing

$$T_\ell E \times T_\ell E \longrightarrow \mathbb{Z}_\ell(1) := \varprojlim \mu_{\ell^n}.$$

Part (b) is proved in [Se5], ch. IV. Part (c) is the main result of [Se6]. □

The following deep result, which is stronger than the second assertion of part (b) of proposition 2.8, is due to Mazur [Maz2], thms. 1 and 2. (Parts (b) and (c) can actually be deduced directly from theorem 1.7, using the fact that a semi-stable curve with a rational subgroup of order ℓ is necessarily isogenous to a curve with a rational point of order ℓ.)

Theorem 2.9 *Suppose that $E/\mathbb{Q}$ is an elliptic curve.*

(a) *If $\ell > 163$ is a prime then $\bar{\rho}_{E,\ell}$ is irreducible.*

(b) *If E is semistable then $\bar{\rho}_{E,\ell}$ is irreducible for $\ell > 7$.*

(c) *If E is semistable and $\bar{\rho}_{E,2}$ is trivial then $\bar{\rho}_{E,\ell}$ is irreducible for $\ell > 3$.*

Remark 2.10 Combined with the results of Serre [Se6], Mazur's results imply that if E is semistable everywhere then $\bar{\rho}_{E,\ell}$ is surjective for $\ell > 7$ ([Maz2], thm. 4).

Local behaviour of $\rho_{E,\ell}$ and $\bar{\rho}_{E,\ell}$:

Proposition 2.11 *Suppose E has good reduction at p.*

(a) *If $\ell \neq p$, then $\rho_{E,\ell}$ is unramified at p, and we have the formula*

$$\operatorname{tr} \rho_{E,\ell}(\mathrm{Frob}_p) = p + 1 - \#\bar{E}_p(\mathbb{F}_p).$$

In particular $\operatorname{tr} \rho_{E,\ell}(\mathrm{Frob}_p)$ belongs to $\mathbb{Z}$ and is independent of $\ell \neq p$.

(b) *For all $n \geq 1$ there is a finite flat group scheme $\mathcal{F}_n/\mathbb{Z}_p$ such that*

$$E[p^n](\bar{\mathbb{Q}}_p) \cong \mathcal{F}_n(\bar{\mathbb{Q}}_p)$$

as G_p-modules.

(c)
- *If E has good ordinary reduction at p (i.e., a_p is not divisible by p) then $\bar{E}_p[p](\bar{\mathbb{F}}_p)$ has order p and*

$$\rho_{E,p}|_{I_p} \sim \begin{pmatrix} \epsilon_p & * \\ 0 & 1 \end{pmatrix};$$

- *If E has supersingular reduction at p, then $\bar{E}_p[p](\bar{\mathbb{F}}_p)$ is trivial and $\bar{\rho}_{E,p}|_{G_p}$ is irreducible.*

Proofs: The proof of part (a) can be found in chapters V and VII of [Sil]. For part (b) one considers the finite flat group scheme $\mathcal{F}_n = \mathcal{E}[p^n]$, where $\mathcal{E}$ is the model for E over $\mathbb{Z}_p$ defined by $(W^{\min})$. (See [Sha] for an introduction to theory of finite flat group schemes.) Part (c) follows from the results in chapters V and VII of [Sil] together with basic results on finite flat group schemes.

Proposition 2.12 *Suppose that E has multiplicative reduction at p. Let $q = q_{E,p} \in p\mathbb{Z}_p$ be the multiplicative Tate period attached to E, defined as in section 1.1. Let $\delta : G_p/I_p \to \{\pm 1\}$ be the unique non-trivial unramified quadratic character of G_p if E has non-split multiplicative reduction, and let δ be the trivial character if E has split multiplicative reduction. Then*

(a)

$$\rho_{E,\ell}|_{G_p} \sim \begin{pmatrix} \epsilon_\ell & * \\ 0 & 1 \end{pmatrix} \otimes \delta,$$

and if $\ell \neq p$ then $m_p(\rho_{E,\ell}) = 1$.

(b)

$$\bar{\rho}_{E,\ell}|_{G_p} \sim \begin{pmatrix} \epsilon_\ell & \Psi \\ 0 & 1 \end{pmatrix} \otimes \delta,$$

and $\Psi \in H^1(G_{\mathbb{Q}_p}, \mathbb{Z}/\ell\mathbb{Z}(1)) \cong \mathbb{Q}_p^\times/(\mathbb{Q}_p^\times)^\ell$ corresponds to the Tate period $q_{E,p} \in \mathbb{Q}_p^\times/(\mathbb{Q}_p^\times)^\ell$.

(c) *If $\ell \neq p$ then $m_p(\bar{\rho}_{E,\ell}) = 0$ (i.e., $\bar{\rho}_{E,\ell}$ is unramified at p) if and only if $\ell | v_p(\Delta_E^{min}) = -v_p(j_E)$.*

(d) *There is a finite flat group scheme $\mathcal{F}/\mathbb{Z}_p$ such that $E[p](\bar{\mathbb{Q}}_p) \cong \mathcal{F}(\bar{\mathbb{Q}}_p)$ if and only if $p | v_p(\Delta_E^{min}) = -v_p(j_E)$.*

Sketch of proof: By Tate's proposition 1.5 of section 1.1, we have

$$E(\bar{\mathbb{Q}}_p) \cong (\bar{\mathbb{Q}}_p^\times / q^{\mathbb{Z}})(\delta)$$

as G_p-modules. (The (δ) indicates that G_p acts on $(\bar{\mathbb{Q}}_p^\times / q^{\mathbb{Z}})(\delta)$ by $\sigma : x \mapsto \sigma(x)^{\delta(\sigma)}$.) The proposition follows directly from this, together with [Edi], prop. 8.2. for the last part. (For further details, see ch. VII, prop. 5.1 of [Si1], ch. IV, thm. 10.2. and ch. V of [Si2], and sec. 2.9 of [Se7].) □

Proposition 2.13 *Suppose that E has additive reduction at p. Then for any $\ell \neq p$ the conductor of $\rho_{E,\ell}|_{G_p}$ is at least 2, and if $p > 3$ then it is equal to 2.*

Remark 2.14 For an elliptic curve E with any type of reduction at p, the conductor of $\rho_{E,\ell}|_{G_p}$ (for $\ell \neq p$) coincides with the local conductor of E at p given by a formula of Ogg (proved by Saito [Sa] in the case $p = 2$). In particular the conductor of $\rho_{E,\ell}|_{G_p}$ is independent of $\ell \neq p$. For further discussion and references, see ch. IV, secs. 10 and 11 of [Si2].

The Frey curve: We now consider the elliptic curve

$$E_{A,B} : Y^2Z = X(X - AZ)(X + BZ),$$

where A and B are non-zero coprime integers with $A+B \neq 0$. Set $C = A+B$. This equation has discriminant $\Delta = 16(ABC)^2$. If p is an odd prime then this curve has good reduction at p if $p \nmid (ABC)$ and has multiplicative reduction at p if $p | ABC$. Thus $\Delta_{E_{A,B}}^{\min} = 2^{4-12n}(ABC)^2$ for some $n \in \mathbb{Z}_{\geq 0}$. If for instance $A \equiv -1 \bmod 4$ and $B \equiv 0 \bmod 16$ then $E_{A,B}$ has a minimal Weierstrass equation

$$Y^2Z + XYZ = X^3 + \frac{B - A - 1}{4} X^2Z - (AB/16)XZ^2.$$

In this special case $E_{A,B}$ has semi-stable reduction everywhere and $\Delta_{E_{A,B}}^{\min} = 2^{-8}(ABC)^2$.

The connection with Fermat's Last Theorem is provided by the elliptic curve

$$E : Y^2Z = X(X - a^\ell Z)(x + b^\ell Z)$$

considered by Hellegouarch [He] and Frey [Fr], where $a^\ell + b^\ell = c^\ell$ is a hypothetical non-trivial solution to Fermat's Last theorem, and where we have supposed without loss of generality that a, b and c are pairwise coprime, that b is even and that $a \equiv -1 \bmod 4$. Frey suggested that properties of this

hypothetical elliptic curve E could lead to a contradiction if the curve were also known to be modular. This suggestion was made precise by Serre [Se7], who observed that the following list of properties would yield the desired contradiction, when combined with his conjectures on Galois representations associated to modular forms (to be discussed in section 3.2).

Theorem 2.15 *Suppose that $\ell > 3$ is prime. The representation $\bar{\rho}_{E,\ell}$ has the following properties.*

(a) *$\bar{\rho}_{E,\ell}$ is irreducible.*

(b) *$\bar{\rho}_{E,\ell}$ is unramified outside ℓ and 2.*

(c) *There is a finite flat group scheme $\mathcal{F}/\mathbb{Z}_\ell$ such that $\mathcal{F}(\bar{\mathbb{Q}}_\ell) \cong E[\ell](\bar{\mathbb{Q}}_\ell)$ as G_ℓ-modules.*

(d) $\#\bar{\rho}_{E,\ell}(I_2) = \ell$.

Proof: Part (a) follows from Mazur's theorem (theorem 2.9, (c)). To prove (b), (c), and (d), we note the key fact that $\Delta_E^{\min} = 2^{-8}(abc)^{2\ell}$ is a perfect ℓ-th power, up to powers of 2. The fact that $\bar{\rho}_{E,\ell}$ is unramified at $p \neq \ell, 2$ (part (b)) follows from this, from proposition 2.11, (a) (for primes $p \neq \ell$ of good reduction for E) and from proposition 2.12 (c) (for primes $p \neq \ell, 2$ of multiplicative reduction for E). Part (c) is a consequence of proposition 2.12 (d), and part (d) follows from proposition 2.12 (c). □

Remark 2.16 Part (a) implies that the image of $\bar{\rho}_{E,\ell}$ is large, while parts (b) and (c) state that the image has very limited ramification. We remark that the conclusion of the theorem ensures that $m_p(\bar{\rho}_{E,\ell}) = 0$ for $p \neq 2, \ell$ and that $m_2(\bar{\rho}_{E,\ell}) = 1$.

2.3 Galois cohomology

In this section M will denote a continuous discrete $G_{\mathbb{Q}}$-module of finite cardinality. If M and N are two such modules we will endow the space $\mathrm{Hom}(M, N)$ of homomorphisms of abelian groups with an action of $G_{\mathbb{Q}}$ via the formula

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

We will use M^* to denote $\mathrm{Hom}(M, \mu_n(\bar{\mathbb{Q}}))$, where n is an integer such that $nM = (0)$ and where $\mu_n(\bar{\mathbb{Q}})$ denotes the group of n^{th} roots of unity in $\bar{\mathbb{Q}}$. We will use $\vee$ to denote Pontryagin duals. All cohomology groups of profinite groups will mean continuous cohomology. The general references for this section are Milne [Mi] and Serre [Se1].

We start by recalling Tate's local duality theorem (see chapter 1, section 2 of [Mi]).

Theorem 2.17 *Let v be a place of $\mathbb{Q}$ (i.e. either a prime number or ∞).*

(a) *$H^i(G_v, M)$ is finite for all i.*

(b) *For all integers n there are compatible embeddings*

$$H^2(G_v, \mu_n(\bar{\mathbb{Q}})) \hookrightarrow \mathbb{Q}/\mathbb{Z}$$

(i.e. compatible with the maps coming from $\mu_n(\bar{\mathbb{Q}}) \hookrightarrow \mu_m(\bar{\mathbb{Q}})$ if $n|m$).

(c) *For $i = 0, 1, 2$ the cup product and the above embeddings give rise to a perfect pairing*

$$H^i(G_v, M) \times H^{2-i}(G_v, M^*) \to H^2(G_v, \mu_n(\bar{\mathbb{Q}})) \hookrightarrow \mathbb{Q}/\mathbb{Z}.$$

(d) *If $v \neq \infty$ is a prime, then $H^i(G_v, M) = (0)$ for $i > 2$ and*

$$\#H^1(G_v, M) = \#H^0(G_v, M)\#H^2(G_v, M)\#(M \otimes \mathbb{Z}_v).$$

(e) *If $v \neq \infty$ and $M \otimes \mathbb{Z}_v = 0$ then $H^1(G_v/I_v, M^{I_v})$ and $H^1(G_v/I_v, M^{*I_v})$ are the annihilators of each other under the pairing*

$$H^1(G_v, M) \times H^1(G_v, M^*) \to \mathbb{Q}/\mathbb{Z}.$$

By a collection of local conditions for M we shall mean a collection $\mathcal{L} = \{L_v\}$ of subgroups $L_v \subset H^1(G_v, M)$ as v runs over all places of $\mathbb{Q}$, such that $L_v = H^1(G_v/I_v, M^{I_v})$ for all but finitely many v. Note that if $\mathcal{L}$ is a collection of local conditions for M then $\mathcal{L}^* = \{L_v^\perp\}$ is a collection of local conditions for M^*. If $\mathcal{L}$ is a collection of local conditions for M we define the corresponding Selmer group, $H^1_{\mathcal{L}}(\mathbb{Q}, M)$, to be the subgroup of $x \in H^1(G_{\mathbb{Q}}, M)$ such that for all places v of $\mathbb{Q}$ we have

$$\mathrm{res}_v x \in L_v \subset H^1(G_v, M).$$

Suppose that v is a finite place of $\mathbb{Q}$. We have an exact sequence

$$(0) \to H^0(G_v, M) \to M^{I_v} \xrightarrow{\mathrm{Frob}_v - 1} M^{I_v} \to H^1(G_v/I_v, M^{I_v}) \to (0),$$

and hence we see that $\#H^1(G_v/I_v, M^{I_v}) = \#H^0(G_v, M)$.

We have the following important observation of Wiles [W3], inspired by a formula of Greenberg [Gre]. An important theme in number theory has been the calculation of Selmer groups. This result, although it does not allow the absolute calculation of Selmer groups, allows the comparison of two (dual) Selmer groups. In the applications in this paper we shall apply the theorem where the various data have been chosen to make one of the Selmer groups trivial. In such a situation it allows the exact calculation of the order of the non-trivial Selmer group.

Theorem 2.18 *If $\mathcal{L}$ is a collection of local conditions for M then the Selmer group $H^1_{\mathcal{L}}(\mathbb{Q}, M)$ is finite. Moreover we have the formula*

$$\frac{\#H^1_{\mathcal{L}}(\mathbb{Q}, M)}{\#H^1_{\mathcal{L}^*}(\mathbb{Q}, M^*)} = \frac{\#H^0(G_{\mathbb{Q}}, M)}{\#H^0(G_{\mathbb{Q}}, M^*)} \prod_v \frac{\#L_v}{\#H^0(G_v, M)}.$$

We note that all but finitely many terms in the product are 1 so that it makes sense.

Proof of theorem 2.18: Choose a finite set, S, of places of $\mathbb{Q}$, which contains ∞, all the places whose residue characteristic divides the order of M, all places at which M is ramified and all places at which $L_v \neq H^1(G_v/I_v, M^{I_v})$. Let $\mathbb{Q}_S$ denote the maximal extension of $\mathbb{Q}$ unramified outside S and let $G_S = \mathrm{Gal}(\mathbb{Q}_S/\mathbb{Q})$. Then we have an exact sequence

$$(0) \to H^1_{\mathcal{L}}(\mathbb{Q}, M) \to H^1(G_S, M) \to \bigoplus_{v \in S} H^1(G_v, M)/L_v.$$

Because $H^1(G_S, M)$ is finite (see cor. 4.15 of ch. 1 of [Mi]), we see that $H^1_{\mathcal{L}}(\mathbb{Q}, M)$ is finite.

Dualising this sequence for M^* and using theorem 2.17, we see that we have an exact sequence

$$\bigoplus_{v \in S} L_v \to H^1(G_S, M^*)^\vee \to H^1_{\mathcal{L}^*}(\mathbb{Q}, M^*)^\vee \to (0).$$

On the other hand we have the following part of the Poitou-Tate nine term sequence (see thm. 4.10 of ch. 1 of [Mi]):

$$\begin{array}{ccccc} (0) \to & H^0(G_S, M) & \to \left(\bigoplus_{v \in S} H^0(G_v, M)\right)/(1+c)M & \to & H^2(G_S, M^*)^\vee \\ & H^1(G_S, M^*)^\vee \leftarrow & \bigoplus_{v \in S} H^1(G_v, M) & \leftarrow & H^1(G_S, M), \end{array}$$

where we regard $(1+c)M$ as contained in $H^0(G_\infty, M)$. Replacing $H^1(G_v, M)$ by L_v for $v \in S$ and combining this with the previous exact sequence we get

$$\begin{array}{ccccc} (0) & \to & H^0(G_S, M) & \to & \left(\bigoplus_{v \in S} H^0(G_v, M)\right)/(1+c)M \\ \bigoplus_{v \in S} L_v & \leftarrow & H^1_{\mathcal{L}}(\mathbb{Q}, M) & \leftarrow & H^2(G_S, M^*)^\vee \\ H^1(G_S, M^*)^\vee & \to & H^1_{\mathcal{L}^*}(\mathbb{Q}, M^*)^\vee & \to & (0). \end{array}$$

Theorem 2.18 follows from this and the global Euler characteristic formula (see thm. 5.1 of ch. 1 of [Mi]):

$$\frac{\#H^1(G_S, M^*)}{\#H^0(G_S, M^*)\#H^2(G_S, M^*)} = \frac{(\#M^*)}{\#H^0(G_\infty, M^*)} = \#(1+c)M.$$

□

We remark, but will not use elsewhere in this article, that the above argument works over any number field:

Theorem 2.19 *Let F be a number field and M a discrete G_F module of finite cardinality. For each place v of F let G_v denote a decomposition group at v and if v is finite let $I_v \subset G_v$ denote the inertia group. Fix subgroups $L_v \subset H^1(G_v, M)$, and such that for all but finitely many v,*

$$L_v = H^1(G_v/I_v, M^{I_v}).$$

Let $H^1_{\mathcal{L}}(F, M)$ (respectively, $H^1_{\mathcal{L}^*}(F, M^*)$) denote the inverse image of $\prod_v L_v$ (respectively, $\prod_v L_v^\perp$) under the map $H^1(G_F, M) \to \prod_v H^1(G_v, M)$ (respectively, $H^1(G_F, M^*) \to \prod_v H^1(G_v, M^*)$). Then $H^1_{\mathcal{L}}(G, M)$ and $H^1_{\mathcal{L}^*}(G, M^*)$ are finite and we have the formula

$$\frac{\#H^1_{\mathcal{L}}(F, M)}{\#H^1_{\mathcal{L}^*}(F, M^*)} = \frac{\#H^0(G_F, M)}{\#H^0(G_F, M^*)} \prod_v \frac{\#L_v}{\#H^0(G_v, M)}.$$

2.4 Representations of $G_{\mathbb{Q}_\ell}$

In this section we assume that ℓ is an **odd** prime. If G is any topological group then by a finite $\mathcal{O}[G]$-module we shall mean a discrete $\mathcal{O}$-module of finite cardinality with a continuous action of G. By a profinite $\mathcal{O}[G]$-module we shall mean an inverse limit of finite $\mathcal{O}[G]$-modules.

If M is a profinite $\mathcal{O}[G_\ell]$-module then we will call M

- **good** if for every discrete quotient M' of M there is a finite flat group scheme $\mathcal{F}/\mathbb{Z}_\ell$ such that $M' \cong F(\bar{\mathbb{Q}}_\ell)$ as $\mathbb{Z}_\ell[G_\ell]$-modules;
- **ordinary** if there is an exact sequence

  $$(0) \to M^{(-1)} \to M \to M^{(0)} \to (0)$$

  of profinite $\mathcal{O}[G_\ell]$-modules such that I_ℓ acts trivially on $M^{(0)}$ and by ϵ on $M^{(-1)}$ (equivalently, if for all $\sigma, \tau \in I_\ell$ we have $(\sigma - \epsilon(\sigma))(\tau - 1) = 0$ on M);
- **semi-stable** if M is either good or ordinary.

Suppose that R is a complete Noetherian local $\mathcal{O}$-algebra with residue field k. We will call a continuous representation $\rho : G_\ell \to GL_2(R)$ good, ordinary or semistable, if

$$\det \rho|_{I_\ell} = \epsilon \tag{2.4.1}$$

and if the underlying profinite $\mathcal{O}[G_\ell]$-module, M_ρ is good, ordinary or semi-stable. We write $\bar{\rho}$ for $\rho \bmod \mathfrak{m}_R$. We record the following consequence of Nakayama's lemma.

Lemma 2.20 *If M_ρ and $\bar{\rho}$ are ordinary, then $M_\rho^{(-1)}$ and $M_\rho^{(0)}$ are each free of rank one over R and ρ is ordinary.*

Remark 2.21 (a) These definitions are somewhat ad hoc, but at the moment that is all that seems to be available (though the work of Fontaine and Laffaille [FL] and its generalisations may well provide a more systematic setting).

(b) For part of the motivation for our definitions, see proposition 2.23 and remark 2.24. For further motivation, we shall see in the next chapter that representations arising from certain modular forms are semistable. Moreover the Fontaine-Mazur conjecture predicts that, conversely, any representation of $G_{\mathbb{Q}}$ with semistable restriction to G_ℓ arises from such a modular form. We shall state a weak form of the Fontaine-Mazur conjecture below (conjecture 3.17).

(c) The terminology in [W3] to describe representations of G_ℓ is slightly different. In particular, we impose the condition (2.4.1) in our definitions of good and ordinary, as this is all we shall need here. Assuming ρ satisfies this condition, the notion of ordinary in [W3] coincides with the one here, and ρ is flat in the sense of [W3] if and only if it is good but not ordinary. (Note that a representation may be both good and ordinary, for instance $\rho_{E,\ell}$ for an elliptic curve with good, ordinary reduction; see proposition 2.23.)

The following assertions, in the good case, are consequences of results of Raynaud [Ray] (in particular, see sec. 2.1, prop. 2.3.1 and cor. 3.3.6 of [Ray]). In the ordinary case they are elementary.

Lemma 2.22 (a) *Good profinite $\mathcal{O}[G_\ell]$-modules are closed under taking sub-objects, quotients and direct products. The same is true for ordinary profinite $\mathcal{O}[G_\ell]$-modules.*

(b) *Suppose that M is a profinite $\mathcal{O}[G_\ell]$-module and that $\{M_i\}$ is a family of sub-objects with trivial intersection, such that each M/M_i is good (resp. ordinary). Then M is good (resp. ordinary).*

(c) *If M is a finite $\mathcal{O}[G_\ell]$-module then M is good if and only if there is a finite flat group scheme $\mathcal{F}/\mathbb{Z}_\ell$ such that $M \cong \mathcal{F}(\bar{\mathbb{Q}}_\ell)$ as $\mathbb{Z}_\ell[G_\ell]$-modules.*

(d) *If M and M' are profinite $\mathcal{O}[G_\ell]$-modules with $M' \cong M$ as $\mathbb{Z}_\ell[I_\ell]$-modules then M is good (resp. ordinary) if and only if M' is good (resp. ordinary).*

(e) *Suppose that M is a profinite $\mathcal{O}[G_\ell]$-module which is finite and free over $\mathcal{O}$. Then M is good if and only if there exists an ℓ-divisible group $\mathcal{F}/\mathbb{Z}_\ell$ such that M is isomorphic to the Tate module of $\mathcal{F}$ as a $\mathbb{Z}_\ell[G_\ell]$-module.*

Together with results stated in section 2.2, we have the following.

Proposition 2.23 *Suppose that E is an elliptic curve over $\mathbb{Q}$ and $\mathcal{O} = \mathbb{Z}_\ell$.*

- *if E has good (resp. semistable) reduction at ℓ, then $\rho_{E,\ell}$ and $\bar{\rho}_{E,\ell}$ are good (resp. semistable);*
- *if E has semistable reduction at ℓ, then $\rho_{E,\ell}$ is ordinary if and only if $\bar{\rho}_{E,\ell}$ is ordinary if and only if either E has multiplicative reduction or E has good ordinary reduction.*

Remark 2.24 It is also true that if $\rho_{E,\ell}$ is good (resp. semistable), then E has good (resp. semistable) reduction at ℓ, but the result is more difficult and we shall not need it.

We will need a few more definitions. We will let $\psi : I_\ell \to \bar{\mathbb{F}}_\ell^\times$ denote the character $\sigma \mapsto (\sigma\varpi)/\varpi \bmod \varpi$ where $\varpi = \sqrt[\ell^2-1]{\ell}$. If F is a field of characteristic other than ℓ and if M is a profinite $\mathbb{Z}_\ell[G_\ell]$-module then we set $M(1) = \varprojlim M \otimes_{\mathbb{Z}_\ell} \mu_{\ell^n}(\bar{F})$. If $\rho : G_\ell \to GL_2(R)$ is ordinary the extension

$$(0) \to R(1) \to M_\rho \to R \to (0)$$

of $R[I_\ell]$-modules gives rise to a class $c_\rho \in H^1(I_\ell, R(1))$. Kummer theory and the valuation on $(\bar{\mathbb{Q}}_\ell^{I_\ell})^\times$ give rise to a map $v : H^1(I_\ell, R(1)) \to H \otimes_{\mathbb{Z}_\ell} R \to R$, where H denotes the ℓ-adic completion of $(\bar{\mathbb{Q}}_\ell^{I_\ell})^\times$. Then we also have the following lemma.

Lemma 2.25 (a) *If $\bar{\rho} : G_\ell \to GL_2(k)$ is good then either $\bar{\rho}$ is ordinary or $\bar{\rho}|_{I_\ell} \otimes \bar{k} = \psi \oplus \psi^\ell$.*

(b) *If $\rho : G_\ell \to GL_2(R)$ is such that M_ρ is good and $\bar{\rho}$ is ordinary, then ρ is good and ordinary.*

(c) *If $\rho : G_\ell \to GL_2(R)$ is ordinary then ρ is good if and only if $v(c_\rho) = 0$.*

Note that we need only consider the case that R has finite cardinality to prove the lemma. Parts (a) and (b) again follow from Raynaud's results [Ray] (for part (b) consider the connected-étale sequences for M_ρ and $M_{\bar{\rho}}$).

We sketch the proof of part (c) using an argument suggested to us by Edixhoven. As in [Edi], prop. 8.2 it suffices to determine which extensions

$$(0) \longrightarrow R(1) \longrightarrow M \longrightarrow R \longrightarrow (0) \tag{2.4.2}$$

of $R[I_\ell]$-modules arise from finite flat group schemes over $\mathbb{Z}_\ell^{\mathrm{nr}}$, the ring of integers of the maximal unramified extension of $\mathbb{Z}_\ell$. By prop. 17.4 of [Oo], the extension (2.4.2) arises from a finite flat group scheme if and only if it corresponds to an extension of sheaves of R-modules in the fpqc topology over $\mathbb{Z}_\ell^{\mathrm{nr}}$. Therefore we must compute the image of

$$\mathrm{Ext}^1_{R-\mathrm{mod}/\mathrm{fpqc}}(R, R(1)) \to \mathrm{Ext}^1_{R[I_\ell]}(R, R(1)).$$

Since the sheaf Ext^1 of R by $R(1)$ vanishes, this is equivalent to computing the image of

$$H^1_{\mathrm{fpqc}}(\mathbb{Z}_\ell^{\mathrm{nr}}, R(1)) \to H^1(I_\ell, R(1)).$$

Part (c) follows from the fact this is precisely the kernel of v.

Remark 2.26 The authors expect that the theory of Fontaine and Laffaille [FL] discussed in the next section could be used to prove that if

$$\rho : G_\ell \to GL_2(R)$$

is such that M_ρ is good and $\bar{\rho}$ is good, then $\det \rho|_{I_\ell}$ is cyclotomic, i.e., ρ is good.

For semistable representations $\rho : G_\ell \to GL_2(\mathcal{O}/\lambda^n)$ we shall define $\mathcal{O}$-submodules

$$H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho) \subset H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}\rho) \subset H^1(G_\ell, \mathrm{ad}\rho).$$

Before doing so, let us consider more generally a continuous representation ρ of a profinite group G with values in $GL_d(\mathcal{O}/\lambda^n)$. For each continuous cocycle $\xi : G \to \mathrm{ad}\rho$, we define a representation

$$\begin{aligned} \rho_\xi : G &\to \quad GL_2(R_n) \\ g &\mapsto (1 + \varepsilon\xi(g))\rho(g) \end{aligned}$$

where $R_n = (\mathcal{O}/\lambda^n)[\varepsilon]/(\varepsilon^2)$. We find that the map $\xi \to \rho_\xi$ induces a bijection between $H^1(G, \mathrm{ad}\rho)$ and equivalence classes of representations $\rho' : G \to GL_2(R_n)$ such that $\rho = \rho' \bmod \varepsilon R_n$, where ρ'_1 and ρ'_2 are deemed equivalent if they are conjugate by an element of $1 + \varepsilon M_2(R_n)$. Now let M_ρ denote the $(\mathcal{O}/\lambda^n)[G]$-module corresponding to ρ, and for each continuous cocycle ξ, let E_ξ denote the $R_n[G]$-module corresponding to ρ_ξ. Note that multiplication by ε defines an isomorphism from $M_\rho = E_\xi/\varepsilon E_\xi$ to εE_ξ of $(\mathcal{O}/\lambda^n)[G]$-modules. We thus obtain an extension

$$0 \to M_\rho \to E_\xi \to M_\rho \to 0$$

and hence a class in $\mathrm{Ext}^1(M_\rho, M_\rho)$ in the category of profinite $(\mathcal{O}/\lambda^n)[G]$-modules. Moreover if ρ_{ξ_1} and ρ_{ξ_2} are equivalent, then E_{ξ_1} and E_{ξ_2} define the same extension class. We thus obtain a map

$$H^1(G, \mathrm{ad}\rho) \to \mathrm{Ext}^1(M_\rho, M_\rho),$$

which the reader can check is an $(\mathcal{O}/\lambda^n)$-linear isomorphism. Classes in $H^1(G, \mathrm{ad}^0\rho)$ correspond to the equivalence classes of representations ρ' satisfying $\det \rho' = \det \rho$, and ϕ in $\mathrm{Hom}(G, \mathcal{O}/\lambda^n)$ corresponds to the twist of ρ by the character $g \mapsto 1 + \varepsilon\phi(g)$. In the case $\ell \nmid d$ where we have the decomposition

$$H^1(G, \mathrm{ad}\rho) = H^1(G, \mathrm{ad}^0\rho) \oplus \mathrm{Hom}(G, \mathcal{O}/\lambda^n),$$

we see that if ρ' corresponds to a class in $H^1(G, \mathrm{ad}\rho)$, then the projection to $\mathrm{Hom}(G, \mathcal{O}/\lambda^n)$ is the homomorphism ϕ such that $\det \rho' = (1 + \varepsilon d\phi) \det \rho$.

We now return to the case of a semistable representation

$$\rho : G_\ell \to GL_2(\mathcal{O}/\lambda^n)$$

and define the cohomology groups $H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho)$ and $H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}\rho)$ as follows. Let $H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}\rho)$ denote the natural image in $H^1(G_\ell, \mathrm{ad}\rho)$ of $\mathrm{Ext}^1(M_\rho, M_\rho)$ taken in the category of semi-stable profinite $\mathcal{O}/\lambda^n[G_\ell]$-modules. If ρ is not good then $H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho)$ is defined to be $H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}\rho)$. If, however, ρ is good then $H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho)$ will denote the natural image in $H^1(G_\ell, \mathrm{ad}\rho)$ of the group $\mathrm{Ext}^1(M_\rho, M_\rho)$ taken in the category of good profinite $\mathcal{O}/\lambda^n[G_\ell]$-modules. We define $H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}^0\rho)$ (resp. $H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}^0\rho)$) as the intersection of $H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho)$ (resp. $H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}\rho)$) and $H^1(G_\ell, \mathrm{ad}^0\rho)$. Note that if ρ is good (resp. semistable) and $\xi : G_\ell \to \mathrm{ad}\rho$ is a cocycle, then E_ξ is good (resp. semistable) if and only if ρ_ξ is.

Suppose that $\rho : G_\ell \to GL_2(\mathcal{O}/\lambda^n)$ is ordinary. Consider the exact sequence

$$0 \to M_\rho^{(-1)} \to M_\rho \to M_\rho^{(0)} \to 0,$$

where $M_\rho^{(-1)}$ denote the maximal submodule of M_ρ where I_ℓ acts by ϵ. Let $\mathrm{ad}^{(-1)}\rho$ denote the submodule $\mathrm{Hom}(M_\rho^{(0)}, M_\rho^{(-1)})$ of $\mathrm{ad}^0\rho$. Then

$$H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}\rho) = \ker\left(H^1(G_\ell, \mathrm{ad}\rho) \to H^1(I_\ell, \mathrm{ad}\rho/\mathrm{ad}^{(-1)}\rho)\right).$$

The same is true with ad^0 replacing ad.

If $\rho : G_\ell \to GL_2(\mathcal{O})$ is semi-stable then we set

$$H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho \otimes K/\mathcal{O}) = \varinjlim H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho \otimes (\lambda^{-n}/\mathcal{O})) \subset H^1(G_\ell, \mathrm{ad}\rho \otimes K/\mathcal{O}).$$

We define $H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}^0\rho \otimes K/\mathcal{O})$ as

$$H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho \otimes K/\mathcal{O}) \cap H^1(G_\ell, \mathrm{ad}^0\rho \otimes K/\mathcal{O}).$$

We make similar definitions for $H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}\rho \otimes K/\mathcal{O})$ and $H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}^0\rho \otimes K/\mathcal{O})$. We will need the following calculations.

Proposition 2.27 (a) *Suppose that $\bar\rho : G_\ell \to GL_2(k)$ is semi-stable. Then*

$$\#H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}^0\bar\rho) \le \#H^0(G_\ell, \mathrm{ad}^0\bar\rho)\#k$$

and equality holds if $\bar\rho$ is ordinary[3].

(b) *Suppose that $\rho : G_\ell \to GL_2(\mathcal{O})$ is both good and ordinary. Let χ_1 and χ_2 be the unramified characters such that $\rho \sim \begin{pmatrix} \chi_1 \epsilon & * \\ 0 & \chi_2 \end{pmatrix}$. Let*

$$c_\ell = (\chi_1/\chi_2)(\mathrm{Frob}_\ell) - 1.$$

[3]The authors expect equality to hold without this hypothesis; cf. remark 2.26.

If $c_\ell \neq 0$, then

$$H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}^0\rho \otimes K/\mathcal{O})/H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}^0\rho \otimes K/\mathcal{O})$$

is finite of order $\#(\mathcal{O}/c_\ell\mathcal{O})$.

(c) *If $\rho \bmod \lambda$ is either not good or not ordinary, then*

$$H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}^0\rho \otimes K/\mathcal{O}) = H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}^0\rho \otimes K/\mathcal{O}).$$

Proof: Part (c) is clear, and for parts (a) and (b) it suffices to prove the following two results.

Proposition 2.28 *Suppose that $\rho : G_\ell \to GL_2(\mathcal{O}/\lambda^n)$ is good. Then*

$$\#H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho) = \#H^0(G_\ell, \mathrm{ad}^0\rho)(\#\mathcal{O}/\lambda^n)^2.$$

Lemma 2.29 *Suppose that $\rho : G_\ell \to GL_2(\mathcal{O}/\lambda^n)$ is ordinary, so that we have $\rho \sim \begin{pmatrix} \chi_1\epsilon & * \\ 0 & \chi_2 \end{pmatrix}$, for some unramified characters χ_1 and χ_2.*

(a) *We have*

$$\#H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}^0\rho) \leq \#H^0(G_\ell, \mathrm{ad}^0\rho)\#(\mathcal{O}/\lambda^n)\#(\mathcal{O}/(\lambda^n, c_\ell))$$

where $c_\ell = (\chi_1/\chi_2)(\mathrm{Frob}_\ell) - 1$. Moreover if ρ is good, then equality holds.

(b) *If $n = 1$ and ρ is not good then $\#H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}^0\rho) = \#k$.*

End of proof of proposition 2.27: To deduce proposition 2.27 (in the good case) from proposition 2.28 and lemma 2.29, note that if $\rho : G_\ell \to GL_2(\mathcal{O}/\lambda^n)$ is good, then

$$H^1(G_\ell/I_\ell, \mathcal{O}/\lambda^n) \subset H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho) \cap H^1(G_\ell, \mathcal{O}/\lambda^n)$$

by lemma 2.22 (d), and this gives the inequality in proposition 2.27(a). Furthermore if ρ is also ordinary, then the above group is contained in

$$\mathrm{Im}\,(H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho) \to H^1(G_\ell, \mathcal{O}/\lambda^n)) \subset H^1(G_\ell/I_\ell, \mathcal{O}/\lambda^n)$$

where the last inclusion comes from lemma 2.25 (b). Therefore

$$\#H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}^0\rho) = \#H^0(G_\ell, \mathrm{ad}^0\rho)\#(\mathcal{O}/\lambda^n).$$

□

We will prove proposition 2.28 in the next section using the theory of Fontaine and Laffaille. The proof of lemma 2.29 is a somewhat technical exercise in the Galois cohomology of local fields, for which we refer the reader to ch. 1 of [W3], prop. 1.9, parts (iii) and (iv). We remark that for our

purposes, inequalities would suffice in part (b) of proposition 2.27 and part (a) of lemma 2.29 (and this is all that is proved in [W3]). We have included the precise formulas for the sake of completeness, since they are not much more difficult to obtain. The only additional observation required is that if ρ is good, then the composite of the natural maps

$$H^1(G_\ell/I_\ell, (\mathrm{ad}^0\rho/\mathrm{ad}^{(-1)}\rho)^{I_\ell}) \to H^1(G_\ell, \mathrm{ad}^0\rho/\mathrm{ad}^{(-1)}\rho) \to H^2(G_\ell, \mathrm{ad}^{(-1)}\rho)$$

is trivial. To prove this, rewrite the composite as

$$H^1(G_\ell/I_\ell, \mathcal{O}/\lambda^n) \to H^1(G_\ell, \mathcal{O}/\lambda^n) \to H^2(G_\ell, \mathrm{ad}^{(-1)}\rho)$$

with the second map given by $\cup c_\rho$ where c_ρ is the class in $H^1(G_\ell, \mathrm{ad}^{(-1)}\rho)$ defining the extension M_ρ and apply lemma 2.25 (c). (In fact, one only needs the easier half of lemma 2.25 (c): if ρ is good then $v(c_\rho) = 0$.)

2.5 The theory of Fontaine and Laffaille

In this section we again assume that ℓ is an odd prime. We mentioned in the last section the importance of understanding good representations of G_ℓ. However the definition of good is somewhat indirect, and this makes computations difficult. The key result we use to address the problem is an equivalence between the category $\mathcal{FF}_\mathcal{O}$ of good finite $\mathcal{O}[G_\ell]$-modules and a category $\mathcal{MF}_\mathcal{O}$ which we define below following Fontaine and Laffaille [FL]. The beauty and utility of the result stems from the elementary algebraic nature of the definition of $\mathcal{MF}_\mathcal{O}$; we can convert questions about good representations into questions in linear algebra.

Remark 2.30 Let $\mathcal{GF}_\mathcal{O}$ denote the category of finite flat commutative group schemes over $\mathbb{Z}_\ell$ with an action of $\mathcal{O}$. By results of Raynaud [Ray], taking $\bar{\mathbb{Q}}_\ell$ points defines an equivalence between the categories $\mathcal{GF}_\mathcal{O}$ and $\mathcal{FF}_\mathcal{O}$. The equivalence between $\mathcal{GF}_\mathcal{O}$ and a category closely related to $\mathcal{MF}_\mathcal{O}$ was first established by Fontaine [Fo2] and [Fo1]. While Fontaine's results would suffice for our purposes here, our formulation will be closer to that in [FL], where an equivalence between $\mathcal{MF}_\mathcal{O}$ and $\mathcal{FF}_\mathcal{O}$ is defined as part of a more general construction of representations of G_ℓ from linear-algebraic data. We caution however that our formulation is not exactly the same as that of [FL] since we wish to work with covariant functors.

We now turn to the definition of $\mathcal{MF}_\mathcal{O}$. The objects are $\mathcal{O}$-modules D of finite cardinality together with a distinguished submodule D^0 and $\mathcal{O}$-linear maps $\phi_{-1} : D \to D$ and $\phi_0 : D^0 \to D$ which satisfy:

- $\phi_{-1}|_{D^0} = \ell\phi_0$,
- $\mathrm{Im}\,\phi_{-1} + \mathrm{Im}\,\phi_0 = D$.

The morphisms are $\mathcal{O}$-linear maps compatible with the additional data of the distinguished submodules and maps ϕ.

It is useful to note that if D is an object of $\mathcal{MF}_{\mathcal{O}}$, then there is a surjection

$$\phi_{-1} \oplus \phi_0 : D/D^0 \oplus D^0/\ell D^0 \twoheadrightarrow D/\ell\phi_0(D^0),$$

and on counting orders we see that this is in fact an isomorphism. Thus there is an isomorphism

$$D/(D^0 + \lambda D) \oplus D^0/\lambda D^0 \xrightarrow{\sim} D/\lambda D.$$

It follows that $D^0/\lambda D^0 \to D/\lambda D$ is injective, and hence that D^0 is (non-canonically) a direct summand of D as an $\mathcal{O}$-module. Note also that ϕ_0 is injective, and if $D = D^0 \oplus D'$ as $\mathcal{O}$-modules, then also $D = \phi_0(D^0) \oplus \phi_{-1}(D')$ as $\mathcal{O}$-modules.

It is then straightforward to check that there is a contravariant functor * from $\mathcal{MF}_{\mathcal{O}}$ to itself defined by:

- $D^* = \mathrm{Hom}(D, \mathbb{Q}_\ell/\mathbb{Z}_\ell)$;
- $(D^*)^0 = \mathrm{Hom}(D/D^0, \mathbb{Q}_\ell/\mathbb{Z}_\ell)$;
- $\phi_{-1}^*(f)(z) = f(\ell x + y)$, where $z = \phi_{-1}(x) + \phi_0(y)$;
- $\phi_0^*(f)(z) = f(x \bmod D^0)$, where $z \equiv \phi_{-1}(x) \bmod (\phi_0 D^0)$.

Moreover the canonical isomorphism $D \cong (D^*)^*$ of $\mathcal{O}$-modules defines a natural isomorphism in $\mathcal{MF}_{\mathcal{O}}$.

We leave it as an exercise for the reader to use the above observations to define cokernels and then kernels of morphisms in $\mathcal{MF}_{\mathcal{O}}$ and verify that it is an abelian category (or see [FL], sec. 1).

Theorem 2.31 *There is a covariant functor $\mathbb{D} : \mathcal{FF}_{\mathcal{O}} \to \mathcal{MF}_{\mathcal{O}}$ which defines an $\mathcal{O}$-additive equivalence of categories. Moreover if M is an object of $\mathcal{FF}_{\mathcal{O}}$, then we have*

(a) *M and $\mathbb{D}(M)$ have the same cardinality;*

(b) *$\mathbb{D}(M) = \mathbb{D}(M)^0$ if and only if M is unramified.*

Remark 2.32 (a) It follows on applying part (a) to $M/\lambda^i M$ for each i that M and $\mathbb{D}(M)$ are in fact (non-canonically) isomorphic as $\mathcal{O}$-modules.

(b) As mentioned above, our formulation differs from that of Fontaine and Laffaille in that we are using covariant functors. To deduce theorem 2.31 from the results in [FL], sec. 9, we define $\mathbb{D}$ as a quasi-inverse of the functor $\underline{U}_S(\cdot^*[1])$, where $\underline{U}_S$ is defined in [FL] and [1] indicates a shift by 1 in filtration degrees. In particular, if $\mathcal{F}$ is a finite flat group scheme over $\mathbb{Z}_\ell$ with an action of $\mathcal{O}$, then the underlying $\mathcal{O}$-module of $\mathbb{D}(\mathcal{F}(\bar{\mathbb{Q}}_\ell))$ can be identified with the covariant Dieudonné module of $\mathcal{F}_{/\mathbf{F}_\ell}$, and ϕ_{-1} with F.

Suppose now that $\rho : G_\ell \to GL_2(\mathcal{O}/\lambda^n)$ is a good continuous representation. Let $D_\rho = \mathbb{D}(M_\rho)$ be the corresponding object of $\mathcal{MF}_\mathcal{O}$. Then $D_\rho \cong (\mathcal{O}/\lambda^n)^2$ as an $\mathcal{O}$-module while $D_\rho^0 \cong \mathcal{O}/\lambda^n$.

Lemma 2.33 *The following are isomorphic.*

(a) *The group of extensions of M_ρ by itself in the category of good finite $\mathcal{O}/\lambda^n[G_\ell]$-modules.*

(b) *The group of extensions of D_ρ by itself in the full subcategory of $\mathcal{MF}_\mathcal{O}$ consisting of objects which are $\mathcal{O}/\lambda^n$-modules.*

(c) *Pairs (α_{-1}, α_0) where $\alpha_{-1} \in \mathrm{Hom}_\mathcal{O}(D_\rho, D_\rho)$, $\alpha_0 \in \mathrm{Hom}_\mathcal{O}(D_\rho^0, D_\rho)$ and $\ell\alpha_0 = \alpha_{-1}|_{D_\rho^0}$, modulo pairs of the form*

$$(a\phi_{\rho,-1} - \phi_{\rho,-1}a, a\phi_{\rho,0} - \phi_{\rho,0}a|_{D^0})$$

where $a \in \mathrm{Hom}_\mathcal{O}(D_\rho, D_\rho)$ and $aD_\rho^0 \subset D_\rho^0$.

Proof: The first two groups of extensions are isomorphic by the Fontaine-Laffaille theorem. Following Ramakrishna [Ram] we explain how to calculate the second group of extensions. We will write D for D_ρ. If

$$(0) \to D \to E \to D \to (0)$$

is an extension then we have as $\mathcal{O}$-modules that $E \cong D^2$ by an isomorphism such that $E^0 \xrightarrow{\sim} (D^0)^2$. Then E is determined by giving elements $\alpha_{E,-1} \in \mathrm{Hom}(D, D)$ and $\alpha_{E,0} \in \mathrm{Hom}(D^0, D)$ with $\ell\alpha_{E,0} = \alpha_{E,-1}|_{D^0}$. Explicitly

$$\phi_{E,-1} = \begin{pmatrix} \phi_{-1} & \alpha_{E,-1} \\ 0 & \phi_{-1} \end{pmatrix} \quad \text{and} \quad \phi_{E,0} = \begin{pmatrix} \phi_0 & \alpha_{E,0} \\ 0 & \phi_0 \end{pmatrix}.$$

Two such extensions E and E' are isomorphic if there is an element $a \in \mathrm{End}\,(D)$ such that $aD^0 \subset D^0$,

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \phi_{-1} & \alpha_{E,-1} \\ 0 & \phi_{-1} \end{pmatrix} = \begin{pmatrix} \phi_{-1} & \alpha_{E',-1} \\ 0 & \phi_{-1} \end{pmatrix} \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix},$$

and

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \phi_0 & \alpha_{E,0} \\ 0 & \phi_0 \end{pmatrix} = \begin{pmatrix} \phi_0 & \alpha_{E',0} \\ 0 & \phi_0 \end{pmatrix} \begin{pmatrix} 1 & a|_{D^0} \\ 0 & 1 \end{pmatrix};$$

or equivalently if $aD^0 \subset D^0$, $\alpha_{E,-1} - \alpha_{E',-1} = \phi_{-1}a - a\phi_{-1}$ and $\alpha_{E,0} - \alpha_{E',0} = \phi_0 a|_{D^0} - a\phi_0$. The lemma follows. □

Corollary 2.34 *The group, $H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho)$, of extensions of M_ρ by itself in the category of good, finite $\mathcal{O}/\lambda^n[G_\ell]$-modules is (non-canonically) isomorphic to $(\mathcal{O}/\lambda^n)^2 \oplus H^0(I_\ell, M_\rho)$.*

Proof: Choose generators e_0, e_{-1} of D, such that e_0 is a generator of D^0. With respect to this basis let ϕ_0 have matrix $\left(\begin{smallmatrix} x \\ z \end{smallmatrix}\right)$ and ϕ_{-1} have matrix $\left(\begin{smallmatrix} \ell x & y \\ \ell z & w \end{smallmatrix}\right)$. If a has matrix $\left(\begin{smallmatrix} a_1 & a_2 \\ 0 & a_3 \end{smallmatrix}\right)$ then a direct calculation shows that

$$(a\phi_{-1} - \phi_{-1}a,\, a\phi_0 - \phi_0 a|_{D^0}) = \left(\begin{pmatrix} \ell a_2 z & (a_1 - a_3)y + a_2(w - \ell x) \\ (a_3 - a_1)\ell z & -\ell z a_2 \end{pmatrix}, \begin{pmatrix} a_2 z \\ (a_3 - a_1)z \end{pmatrix}\right).$$

Thus the Ext group we want is the quotient of the set of pairs of matrices (α_{-1}, α_0) as above by the submodule generated by

$$\left(\begin{pmatrix} \ell z & w - \ell x \\ 0 & -\ell z \end{pmatrix}, \begin{pmatrix} z \\ 0 \end{pmatrix}\right) \quad \text{and} \quad \left(\begin{pmatrix} 0 & -y \\ \ell z & 0 \end{pmatrix}, \begin{pmatrix} 0 \\ z \end{pmatrix}\right).$$

Note that either z or w is a unit in $\mathcal{O}/\lambda$ from which it follows that the Ext group we want is isomorphic to $(\mathcal{O}/\lambda^n)^2 \oplus \mathcal{O}/(z, \lambda^n)$. On the other hand $H^0(I_\ell, M_\rho)$ corresponds to the largest submodule $C \subset D^0$ such that $\phi_0 C = C$, i.e. $C = \{d \in D^0 : zd = 0\} \cong \mathcal{O}/(z, \lambda^n)$. □

Corollary 2.35 *Suppose that $\rho : G_\ell \to GL_2(\mathcal{O}/\lambda^n)$ is a continuous good representation. Then $H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}\rho)$ is isomorphic to $(\mathcal{O}/\lambda^n)^2 \oplus H^0(G_\ell, \mathrm{ad}^0\rho)$.*

Proof: If ρ is not ordinary then $H^0(G_\ell, \mathrm{ad}^0\rho)$ and $H^0(I_\ell, M_\rho)$ are both trivial, so suppose that ρ is ordinary. In this case, let ρ' denote the ordinary representation defined by

$$M_{\rho'} = \mathrm{Hom}(M_\rho^{(0)}, M_\rho) \subset \mathrm{ad}^0(\rho),$$

and let $M_{\rho'}^{(-1)} = \mathrm{ad}^0(\rho)^{(-1)}$. Then

$$H^0(G_\ell, \mathrm{ad}^0\rho) = H^0(G_\ell, M_{\rho'}) \subset H^0(I_\ell, M_{\rho'}) \cong H^0(I_\ell, M_\rho).$$

Since $H^0(I_\ell, M_{\rho'}^{(-1)})$ is trivial, the restriction homomorphism

$$H^1(G_\ell, M_{\rho'}^{(-1)}) \to H^1(I_\ell, M_{\rho'}^{(-1)})$$

is injective. We deduce from the long exact sequences associated to

$$(0) \to M_{\rho'}^{(-1)} \to M_{\rho'} \to \mathcal{O}/\lambda^n \to 0$$

that $H^0(G_\ell, M_{\rho'}) = H^0(I_\ell, M_{\rho'})$. □

2.6 Deformations of representations

In this section we shall review Mazur's theory of deformations of representations of profinite groups (see [Maz3]).

Let $\mathcal{C}_{\mathcal{O}}$ denote the category whose objects are complete noetherian local $\mathcal{O}$-algebras with residue field k and whose morphisms are $\mathcal{O}$-algebra homomorphisms which are local (i.e. take maximal ideals into maximal ideals). (The structure maps from $\mathcal{O}$ to every object of $\mathcal{C}_{\mathcal{O}}$ are also assumed to be local.) Let G denote a topologically finitely generated profinite group and let $\bar{\rho}$ denote an absolutely irreducible representation of G into $GL_d(k)$. (In fact all we shall use in the sequel is that k is the centraliser in $M_d(k)$ of the image of $\bar{\rho}$.) Let $\mathcal{D}_0$ denote the category of profinite $\mathcal{O}[G]$-modules with continuous morphisms. We will let $\mathcal{D}$ denote a full subcategory of $\mathcal{D}_0$ which is closed under taking sub-objects, quotients and direct products and which contains $M_{\bar{\rho}}$. Note that if M is an object of $\mathcal{D}_0$ and M_i is a collection of subobjects which have trivial intersection and such that each M/M_i is an object of $\mathcal{D}$, then M is an object of $\mathcal{D}$, since $M \subset \prod_i M/M_i$.

Let $\chi : G \to \mathcal{O}^\times$ be a continuous character such that $\det \bar{\rho} = \chi \bmod \lambda$. By a lifting of $\bar{\rho}$ of type $D = (\mathcal{O}, \chi, \mathcal{D})$ we shall mean an object R of $\mathcal{C}_{\mathcal{O}}$ and a continuous representation $\rho : G \to GL_d(R)$ such that:

(a) $\rho \bmod \mathfrak{m}_R = \bar{\rho}$,

(b) $\det \rho = \chi$,

(c) M_ρ is an object of $\mathcal{D}$.

Note that if $\phi : R \to R'$ and $\rho : G \to GL_d(R)$ is of type D so is $\phi \circ \rho$.

Theorem 2.36 *There is a lifting $\rho_D^{\mathrm{univ}} : G \to GL_d(R_D)$ of $\bar{\rho}$ of type D such that if $\rho : G \to GL_d(R)$ is any lifting of $\bar{\rho}$ of type D then there is a unique homomorphism of $\mathcal{O}$-algebras $\phi : R_D \to R$ such that ρ is conjugate to $\phi \circ \rho_D^{\mathrm{univ}}$.*

The representation ρ_D^{univ} is referred to as the *universal deformation* of type D. Mazur [Maz3] proved this theorem for $\mathcal{D}_0$ and certain other categories $\mathcal{D}$. Ramakrishna [Ram] observed that the arguments work with any category $\mathcal{D}$ satisfying the above hypotheses. We will sketch a proof which was suggested by Faltings. (We remark that another explicit construction of the deformation ring has been given by de Smit and Lenstra in [dSL].)

Proof of theorem 2.36: Choose a sequence $g_1, ..., g_r$ of topological generators of G and liftings $A_1, ..., A_r$ of $\bar{\rho}(g_1), ..., \bar{\rho}(g_r)$ to $M_d(\mathcal{O})$. Define a mapping $\iota : M_d(\mathcal{O}) \to M_d(\mathcal{O})^r$ which sends x to $(xA_1 - A_1x, ..., xA_r - A_rx)$. Since ι has torsion-free cokernel, so we can decompose

$$M_d(\mathcal{O})^r = \iota(M_d(\mathcal{O})) \oplus V,$$

for some submodule $V \subset M_d(\mathcal{O})^r$. If $\rho : G \to GL_d(R)$ is a lifting of $\bar{\rho}$ of type D set $v_\rho = (\rho(g_1) - A_1, ..., \rho(g_r) - A_r) \in M_d(R)^r$. Note that $v_\rho \equiv 0 \bmod \mathfrak{m}_R$ and that v_ρ completely determines ρ. We call the lifting ρ well-placed if v_ρ belongs to $V \otimes_{\mathcal{O}} R \subset M_d(R)^r$. The crucial observation is the following result.

Lemma 2.37 *If $\rho : G \to GL_d(R)$ is a lifting of $\bar{\rho}$ of type D then there is a unique conjugate ρ' of ρ which is well-placed.*

The lemma is first proved for algebras R such that $\mathfrak{m}_R^n = (0)$ by induction on n, and then one deduces the general case.

In virtue of the lemma it suffices to find a universal well-placed lifting of type D. Let $e_1, ..., e_s$ be a basis of V as an $\mathcal{O}$-module. If ρ is a well-placed lifting of $\bar{\rho}$ of type D then we can write $v_\rho = \sum_{i=1}^s v_{\rho,i} e_i$ for unique elements $v_{\rho,i} \in \mathfrak{m}_R$ and we can define a homomorphism

$$\theta_\rho : \mathcal{O}[[T_1, ..., T_s]] \to R$$

sending T_i to $v_{\rho,i}$. Note that ρ is completely determined by θ_ρ ($\rho(g_i) = A_i + \sum_{j=1}^s \theta_\rho(T_j) e_{ji}$, where $e_j = (e_{j1}, ..., e_{jr})$). Let I denote the intersection of all ideals J of $\mathcal{O}[[T_1, ..., T_s]]$ such that there is a representation $\rho_J : G \to GL_d(\mathcal{O}[[T_1, ..., T_s]]/J)$ of type D with $\rho_J(g_i) = A_i + \sum_{j=1}^s T_j e_{ji}$ for all i. Let R_D denote the quotient of $\mathcal{O}[[T_1, ..., T_s]]$ by I. Then one can check that there is a representation $\rho^{\text{univ}} : G \to GL_d(R_D)$ with $\rho^{\text{univ}}(g_i) = A_i + \sum_{j=1}^s T_j e_{ji}$ for all i, and that this is the desired universal representation. □

We will need a few elementary properties of these universal deformations. More precisely we will need to know how these universal rings change when we change the base field, we will need to know how to calculate the equicharacteristic tangent space of these rings and more generally how to calculate $\wp/\wp^2$ for certain prime ideals $\wp$. The first of these lemmas is a remark of Faltings, the second is due to Mazur [Maz3] and the third to Wiles [W3].

Lemma 2.38 *Let K'/K be a finite extension with ring of integers $\mathcal{O}'$ and residue field k'. Let $\mathcal{D}'$ denote the full sub-category of the category of profinite $\mathcal{O}'[G]$-modules such that the underlying object of $\mathcal{D}_0$ is actually an object of $\mathcal{D}$. Let $D' = (\mathcal{O}', \chi, \mathcal{D}')$. Then $R_{D'} = R_D \otimes_{\mathcal{O}} \mathcal{O}'$ and $\rho_{D'}^{\text{univ}} = \rho_D^{\text{univ}} \otimes 1$.*

Proof: Let $\tilde{R}_{D'}$ denote the subring of $R_{D'}$ consisting of elements which reduce modulo the maximal ideal to an element of $k \subset k'$. Similarly let $\tilde{R}_D$ denote the subring of $R_D \otimes_{\mathcal{O}} \mathcal{O}'$ consisting of elements which reduce modulo the maximal ideal to an element of $k \subset k'$. Then $\rho_{D'}^{\text{univ}}$ is in fact valued in $GL_d(\tilde{R}_{D'})$ and $\rho_D^{\text{univ}} \otimes 1$ is in fact valued in $GL_d(\tilde{R}_D)$. The universal properties give natural maps $\alpha : R_D \to \tilde{R}_{D'}$ and $\beta : R_{D'} \to R_D \otimes_{\mathcal{O}} \mathcal{O}'$. Moreover they show that the composite $(\alpha \otimes 1) \circ \beta : R_{D'} \to R_{D'}$ is the identity and that $\beta \circ \alpha : R_D \to \tilde{R}_D$ is the natural embedding. Thus β is an isomorphism. □

We will let $\mathcal{D}^{(n)}$ denote the full subcategory of $\mathcal{D}$ whose objects are killed by λ^n. Suppose that M is an object of $\mathcal{D}_0$ which is finite and free over $\mathcal{O}/\lambda^n$. Recall from section 2.4 that $H^1(G, \text{End}(M))$ may be identified with $\text{Ext}^1_{\mathcal{D}_0^{(n)}}(M, M)$. If M is an object of $\mathcal{D}^{(n)}$ which is finite and free over $\mathcal{O}/\lambda^n$, then we have a natural inclusion

$$\text{Ext}^1_{\mathcal{D}^{(n)}}(M, M) \subset \text{Ext}^1_{\mathcal{D}_0^{(n)}}(M, M) \cong H^1(G, \text{End}(M)).$$

We define $H^1_{\mathcal{D}}(G, \mathrm{End}\,(M))$ to be the image of $\mathrm{Ext}^1_{\mathcal{D}^{(n)}}(M, M)$ in the group $H^1(G, \mathrm{End}\,(M))$, and $H^1_{\mathcal{D}}(G, \mathrm{End}^0(M))$ to be the intersection

$$H^1(G, \mathrm{End}^0(M)) \cap H^1_{\mathcal{D}}(G, \mathrm{End}\,(M)).$$

Lemma 2.39 *There is a canonical isomorphism of k-vector spaces*

$$\mathrm{Hom}_k(\mathfrak{m}_{R_D}/(\lambda, \mathfrak{m}^2_{R_D}), k) \cong H^1_{\mathcal{D}}(G, \mathrm{ad}^0\bar{\rho}).$$

Proof: There is a natural bijection between $\mathrm{Hom}_k(\mathfrak{m}_{R_D}/(\lambda, \mathfrak{m}^2_{R_D}), k)$ and the set of $\mathcal{O}$-algebra homomorphisms from R_D to the algebra $k[\varepsilon]$ where $\varepsilon^2 = 0$ (the correspondence associates $\phi : R_D \to k[\varepsilon]$ to $\phi|_{\mathfrak{m}_{R_D}}$). Hence there is a bijection to the set of liftings $\rho : G \to GL_d(k[\varepsilon])$ of $\bar{\rho}$ of type D, modulo conjugation by elements of $1 + \varepsilon M_d(k)$. On the other hand, recall from section 2.4 that there is a natural bijection between $\mathrm{Ext}^1_{\mathcal{D}^{(1)}_0}(M_{\bar{\rho}}, M_{\bar{\rho}})$ and the set of of all continuous liftings $\rho : G \to GL_d(k[\varepsilon])$ of $\bar{\rho}$ modulo conjugation by elements of $1 + \varepsilon M_d(k)$. Moreover a lifting ρ is type D if and only if $\det\rho = \det\bar{\rho}$ and the corresponding extension M_ρ is an object of $\mathcal{D}$. The lemma follows on checking linearity. □

Now suppose that $\theta : R_D \to \mathcal{O}$ is an $\mathcal{O}$-algebra homomorphism. Let $\wp$ denote the kernel of θ and let $\rho = \theta \circ \rho^{\mathrm{univ}}_D$. We set

$$H^1_{\mathcal{D}}(G, \mathrm{ad}^0\rho \otimes K/\mathcal{O}) = \varinjlim H^1_{\mathcal{D}}(G, \mathrm{ad}^0\rho \otimes \lambda^{-n}/\mathcal{O}) \subset H^1(G, \mathrm{ad}^0\rho \otimes K/\mathcal{O}).$$

Lemma 2.40 *There is a canonical $\mathcal{O}$-linear isomorphism*

$$\mathrm{Hom}_{\mathcal{O}}(\wp/\wp^2, K/\mathcal{O}) \cong H^1_{\mathcal{D}}(G, \mathrm{ad}^0\rho \otimes K/\mathcal{O}).$$

Proof: One shows in a very similar manner to the proof of lemma 2.39 that for all n there is a natural isomorphism.

$$\mathrm{Hom}_{\mathcal{O}}(\wp/\wp^2, \mathcal{O}/\lambda^n) \cong H^1_{\mathcal{D}}(G, \mathrm{ad}^0\rho \otimes \mathcal{O}/\lambda^n).$$

□

2.7 Deformations of Galois representations

Again in this section we assume that ℓ is an odd prime. Let $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(k)$ denote a continuous absolutely irreducible representation. Suppose moreover that $\det\bar{\rho} = \epsilon$ and that $\bar{\rho}$ is semi-stable in the sense that

- $\bar{\rho}|_{G_\ell}$ is semi-stable,
- and if $p \neq \ell$ then $\#\bar{\rho}(I_p)|\ell$.

Note that if $E/\mathbb{Q}$ is a semistable elliptic curve then $\bar{\rho}_{E,\ell} : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_\ell)$, satisfies these conditions if it is irreducible. By a theorem of Mazur (theorem 2.9) this will be the case if $\ell > 7$.

Let Σ denote a finite set of prime numbers. If R is an object of $\mathcal{C}_{\mathcal{O}}$ then we say that a continuous lifting $\rho : G_{\mathbb{Q}} \to GL_2(R)$ of $\bar{\rho}$ is of type Σ if the following hold.

- $\det \rho = \epsilon$.
- $\rho|_{G_\ell}$ is semi-stable.
- If $\ell \notin \Sigma$ and $\bar{\rho}|_{G_\ell}$ is good then $\rho|_{G_\ell}$ is good.
- If $p \notin \Sigma \cup \{\ell\}$ and $\bar{\rho}$ is unramified at p then ρ is unramified at p.
- If $p \notin \Sigma \cup \{\ell\}$ then $\rho|_{I_p} \sim \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$.

Roughly speaking we require that at primes $p \notin \Sigma$, ρ is as unramified as could be hoped for and we require that $\rho|_{G_\ell}$ is semi-stable. Note that if $\Sigma \subset \Sigma'$ and ρ is a lifting of type Σ then it is also a lifting of type Σ'. Note also that if $E/\mathbb{Q}$ is an elliptic curve which is semi-stable at ℓ and for which $\bar{\rho}_{E,\ell}$ is irreducible and semi-stable, then $\rho_{E,\ell} : G_{\mathbb{Q}} \to GL_2(\mathbb{Z}_\ell)$ is a lifting of type Σ if Σ contains all the primes for which E has bad reduction.

Now suppose that $\rho : G_{\mathbb{Q}} \to GL_2(\mathcal{O}/\lambda^n)$ is a lifting of $\bar{\rho}$ of type Σ. We will write $H^1_\Sigma(\mathbb{Q}, \mathrm{ad}^0\rho)$ for $H^1_{\mathcal{L}_\Sigma}(\mathbb{Q}, \mathrm{ad}^0\rho)$, where

- $L_{\Sigma,p} = H^1(G_p/I_p, (\mathrm{ad}^0\rho)^{I_p})$ if $p \notin \Sigma \cup \{\ell\}$;
- $L_{\Sigma,p} = H^1(G_p, (\mathrm{ad}^0\rho))$ if $p \in \Sigma$ and $p \neq \ell$;
- $L_{\Sigma,\ell} = H^1_{\mathrm{f}}(G_\ell, (\mathrm{ad}^0\rho))$ if $\ell \notin \Sigma$;
- $L_{\Sigma,\ell} = H^1_{\mathrm{ss}}(G_\ell, (\mathrm{ad}^0\rho))$ if $\ell \in \Sigma$.

Note that the pairing $\mathrm{ad}^0\rho \times \mathrm{ad}^0\rho \to \mathcal{O}/\lambda^n$ given by $(a, b) \mapsto \mathrm{tr}\,(ab)$ is perfect and respects the action of $G_{\mathbb{Q}}$ (i.e. $\mathrm{tr}\,((\mathrm{ad}^0\rho)(g)a, (\mathrm{ad}^0\rho)(g)b) = \mathrm{tr}\,(ab)$ for all $g \in G_{\mathbb{Q}}$). Choosing a generator for the $\mathcal{O}$-module $\mathrm{Hom}(\mathcal{O}/\lambda^n, \mathbb{Q}_\ell/\mathbb{Z}_\ell)$, we obtain an isomorphism of $\mathcal{O}[G_{\mathbb{Q}}]$-modules

$$\mathrm{ad}^0\rho(1) \cong \mathrm{Hom}_{\mathcal{O}}(\mathrm{ad}^0\rho, \mathcal{O}/\lambda^n)(1) \xrightarrow{\sim} \mathrm{Hom}(\mathrm{ad}^0\rho, \mathbb{Q}_\ell/\mathbb{Z}_\ell)(1) = (\mathrm{ad}^0\rho)^*.$$

The $\mathcal{O}$-submodule of $H^1(\mathbb{Q}, \mathrm{ad}^0\rho(1))$ corresponding to $H^1_{\mathcal{L}^*_\Sigma}(\mathbb{Q}, (\mathrm{ad}^0\rho)^*)$ is independent of the choice of generator and we denote it $H^1_\Sigma(\mathbb{Q}, \mathrm{ad}^0\rho(1))$. Thus $H^1_\Sigma(\mathbb{Q}, \mathrm{ad}^0\rho(1))$ is defined by the local conditions $\{L^\perp_{\Sigma,v}\}$ where the orthogonality is with respect to the pairing

$$H^1(G_v, \mathrm{ad}^0\rho) \times H^1(G_v, \mathrm{ad}^0\rho(1)) \to \mathbb{Q}_\ell/\mathbb{Z}_\ell$$

arising from the above isomorphism. We may equivalently regard the orthogonality as being with respect to the natural perfect $\mathcal{O}$-bilinear pairing defined by the composition

$$\begin{aligned} H^1(G_v, \mathrm{ad}^0\rho) \times H^1(G_v, \mathrm{ad}^0\rho(1)) &\to H^2(G_v, \mathrm{ad}^0\rho \otimes_{\mathcal{O}} \mathrm{ad}^0\rho(1)) \\ &\to H^2(G_v, \mathcal{O}/\lambda^n(1)) \cong \mathcal{O}/\lambda^n. \end{aligned}$$

Note that if $p \neq \ell$ we have that

- $L^{\perp}_{\Sigma,p} = H^1(G_p/I_p, (\mathrm{ad}^0\rho)(1)^{I_p})$ if $p \notin \Sigma$;
- $L^{\perp}_{\Sigma,p} = (0)$ if $p \in \Sigma$.

If $\rho : G_{\mathbb{Q}} \to GL_2(\mathcal{O})$ is a lifting of $\bar{\rho}$ of type Σ, we will write $H^1_{\Sigma}(\mathbb{Q}, \mathrm{ad}^0\rho \otimes K/\mathcal{O})$ for the direct limit $\varinjlim H^1_{\Sigma}(\mathbb{Q}, \mathrm{ad}^0\rho \otimes \lambda^{-n}/\mathcal{O})$, and $H^1_{\Sigma}(\mathbb{Q}, \mathrm{ad}^0\rho(1) \otimes K/\mathcal{O})$ for the direct limit $\varinjlim H^1_{\Sigma}(\mathbb{Q}, \mathrm{ad}^0\rho(1) \otimes \lambda^{-n}/\mathcal{O})$.

The main result we need about deformations of $\bar{\rho}$ is the following.

Theorem 2.41 *There is a universal lifting $\rho^{\mathrm{univ}}_{\Sigma} : G_{\mathbb{Q}} \to GL_2(R_{\Sigma})$ of $\bar{\rho}$ of type Σ, i.e. $\rho^{\mathrm{univ}}_{\Sigma}$ is a lifting of type Σ and if $\rho : G_{\mathbb{Q}} \to GL_2(R)$ is any lifting of type Σ then there is a unique $\mathcal{O}$-algebra homomorphism $\phi : R_{\Sigma} \to R$ such that $\rho \sim \phi \circ \rho^{\mathrm{univ}}_{\Sigma}$. Moreover we have the following.*

(a) *If K'/K is a finite extension and R'_{Σ} is the corresponding deformation ring then $R'_{\Sigma} = R_{\Sigma} \otimes_{\mathcal{O}} \mathcal{O}'$ and $(\rho^{\mathrm{univ}}_{\Sigma})' = \rho^{\mathrm{univ}}_{\Sigma} \otimes 1$.*

(b) *R_{Σ} can be topologically generated as an $\mathcal{O}$-algebra by $\dim_k H^1_{\Sigma}(\mathbb{Q}, \mathrm{ad}^0\bar{\rho})$ elements.*

(c) *If $\phi : R_{\Sigma} \twoheadrightarrow \mathcal{O}$ is a $\mathcal{O}$-algebra homomorphism, if $\rho = \phi \circ \rho_{\Sigma}$ and if $\wp = \ker\phi$ then $\mathrm{Hom}(\wp/\wp^2, K/\mathcal{O}) \cong H^1_{\Sigma}(\mathbb{Q}, \mathrm{ad}^0\rho \otimes K/\mathcal{O})$.*

Proof: Let L_0 denote the fixed field of $\bar{\rho}$. Let L_n (for $n \in \mathbb{Z}_{>0}$) denote the maximal elementary abelian ℓ-extension of L_{n-1} unramified outside Σ, $\{\ell\}$ and the primes where $\bar{\rho}$ ramifies. Let $L_{\infty} = \bigcup_n L_n$ and let $G = \mathrm{Gal}\,(L_{\infty}/\mathbb{Q})$. Note that any lifting of $\bar{\rho}$ of type Σ factors through G. $\mathrm{Gal}\,(L_{\infty}/L_0)$ is a pro-ℓ-group and its maximal elementary abelian quotient, $\mathrm{Gal}\,(L_1/L_0)$, is finite by theorem 2.2. We deduce from the following lemma that $\mathrm{Gal}\,(L_{\infty}/L_0)$ and hence G are topologically finitely generated. (See for instance [Koc] Satz 4.10 for a proof of this lemma.)

Lemma 2.42 *Let H be a pro-ℓ-group and $\bar{H}$ its maximal elementary abelian quotient. Suppose $h_1, ..., h_r \in H$ map to a set of generators of $\bar{H}$, then $h_1, ..., h_r$ topologically generate H.*

Let $\mathcal{D}$ denote the category of profinite $\mathcal{O}[G]$-modules M for which

- M is semi-stable as an $\mathcal{O}[G_{\ell}]$-module,
- if $\ell \notin \Sigma$ and if $\bar{\rho}$ is good then M is good as an $\mathcal{O}[G_{\ell}]$-module,
- if $p \notin \Sigma \cup \{\ell\}$ and if $\bar{\rho}$ is ramified at p then there is an exact sequence of $\mathcal{O}[I_p]$-modules

$$(0) \to M^{(-1)} \to M \to M^{(0)} \to (0),$$

such that I_p acts trivially on $M^{(-1)}$ and $M^{(0)}$.

Then we see that a lifting $\rho : G_{\mathbb{Q}} \to GL_2(R)$ of $\bar{\rho}$ is of type Σ if and only if

- ρ factors through G,
- $\det \rho = \epsilon$,
- M_ρ is an object of $\mathcal{D}$.

The existence part of theorem 2.41 now follows from theorem 2.36 and part (a) follows from lemma 2.38. Parts (b) and (c) follow from lemmas 2.39 and 2.40, and the following observation:
If $\rho : G_{\mathbb{Q}} \to GL_2(\mathcal{O}/\lambda^n)$ is a lifting of type Σ and if $p \notin \Sigma \cup \{l\}$ is a prime where $\bar{\rho}$ ramifies then

$$\ker\left(H^1(G_p, \mathrm{ad}^0\rho) \to H^1(I_p, \mathrm{ad}^0\rho/(\mathrm{ad}^0\rho)^{I_p})\right) = H^1(G_p/I_p, (\mathrm{ad}^0\rho)^{I_p}). \tag{2.7.1}$$

Equation (2.7.1) follows from the fact that the natural map

$$H^1(I_p, \mathrm{ad}^0\rho) \to H^1(I_p, \mathrm{ad}^0\rho/(\mathrm{ad}^0\rho)^{I_p})$$

is an injection (in fact an isomorphism). It is nothing other than the map

$$(\mathrm{ad}^0\rho)_{I_p} \to (\mathrm{ad}^0\rho/(\mathrm{ad}^0\rho)^{I_p})_{I_p}.$$

This completes the proof of theorem 2.41. □

Corollary 2.43 *Suppose that if $\ell = 3$ then $\bar{\rho}|_{G_{\mathbb{Q}(\sqrt{-3})}}$ is absolutely irreducible. Then R_Σ can be topologically generated as an $\mathcal{O}$-algebra by*

$$\dim_k H^1_\Sigma(\mathbb{Q}, \mathrm{ad}^0\bar{\rho}(1)) + d_\ell + \sum_{p \in \Sigma - \{\ell\}} \dim_k H^0(\mathbb{Q}_p, \mathrm{ad}^0\bar{\rho}(1))$$

elements, where $d_\ell = \dim_k H^1_{\mathrm{ss}}(\mathbb{Q}_\ell, \mathrm{ad}^0\bar{\rho}) - \dim_k H^1_{\mathrm{f}}(\mathbb{Q}_\ell, \mathrm{ad}^0\bar{\rho})$ if $\ell \in \Sigma$, while $d_\ell = 0$ if $\ell \notin \Sigma$.

Proof. Note that $H^0(\mathbb{Q}, \mathrm{ad}^0\bar{\rho}(1)) = 0$ unless $\ell = 3$ and $\bar{\rho}|_{G_{\mathbb{Q}(\sqrt{-3})}}$ is not absolutely irreducible. Thus according to theorem 2.18, $\dim_k H^1_\Sigma(\mathbb{Q}, \mathrm{ad}^0\bar{\rho})$ is the sum of the following terms.

- $\dim_k H^1_\Sigma(\mathbb{Q}, \mathrm{ad}^0\bar{\rho}(1))$.
- $\dim_k H^1_{\mathrm{f}}(\mathbb{Q}_\ell, \mathrm{ad}^0\bar{\rho}) - \dim_k H^0(\mathbb{Q}_\ell, \mathrm{ad}^0\bar{\rho}) - \dim_k H^0(\mathbb{R}, \mathrm{ad}^0\bar{\rho}) \leq 0$ by proposition 2.27.
- d_ℓ.
- For each $p \in \Sigma - \{l\}$,

  $$\dim_k H^1(\mathbb{Q}_p, \mathrm{ad}^0\bar{\rho}) - \dim_k H^0(\mathbb{Q}_p, \mathrm{ad}^0\bar{\rho}) = \dim_k H^0(\mathbb{Q}_p, \mathrm{ad}^0\bar{\rho}(1))$$

  by the local Euler characteristic formula (see theorem 2.17).

□

2.8 Special cases

In this section we will restrict attention to the deformation problems associated with sets Q of certain special primes q. These sets Q are chosen so that the associated deformation ring R_Q enjoys a number of special properties which will be crucial in the sequel. In particular we will only consider sets Q of primes q which satisfy the following two properties

- $q \equiv 1 \bmod \ell$,
- $\bar{\rho}$ is unramified at q and $\bar{\rho}(\mathrm{Frob}_q)$ has distinct k-rational eigenvalues.

For such primes q we will let Δ_q denote the maximal quotient of $(\mathbb{Z}/q\mathbb{Z})^\times$ of ℓ-power order. Then Δ_q is naturally a quotient of both G_q and $G_{\mathbb{Q}}$ via the map

$$\chi_q : G_{\mathbb{Q}} \twoheadrightarrow \mathrm{Gal}(\mathbb{Q}(\zeta_q)/\mathbb{Q}) \cong (\mathbb{Z}/q\mathbb{Z})^\times \twoheadrightarrow \Delta_q.$$

We will let $\Delta_Q = \prod_{q \in Q} \Delta_q$, $\chi_Q = \prod_{q \in Q} \chi_q : G_{\mathbb{Q}} \twoheadrightarrow \Delta_Q$ and $\mathfrak{a}_Q$ denote the augmentation ideal of $\mathcal{O}[\Delta_Q]$. Note that there is an isomorphism

$$\mathcal{O}[\Delta_Q] \cong \mathcal{O}[[S_q : q \in Q]]/((1+S_q)^{\#\Delta_q} - 1 : q \in Q)$$

under which $\mathfrak{a}_Q$ corresponds to $(S_q : q \in Q)$.

For each $q \in Q$, we choose an eigenvalue α_q of $\bar{\rho}(\mathrm{Frob}_q)$ and denote the other β_q.

Lemma 2.44 *If $q \in Q$ then $\rho_Q^{\mathrm{univ}}|_{G_q}$ is conjugate to $\begin{pmatrix} \xi & 0 \\ 0 & \epsilon\xi^{-1} \end{pmatrix}$ for some character ξ such that $\bar{\xi}(\mathrm{Frob}_q) = \alpha_q$.*

Proof: Choose a lifting $f \in G_q$ of Frob_q. As $\bar{\rho}(f)$ has distinct k-rational eigenvalues it is a simple application of Hensel's lemma to see that we can choose a basis such that $\rho_Q^{\mathrm{univ}}(f) = \begin{pmatrix} \tilde{\alpha}_q & 0 \\ 0 & \tilde{\beta}_q \end{pmatrix}$ where $\tilde{\alpha}_q$ and $\tilde{\beta}_q$ reduce to α_q and β_q in k. It suffices to show that for any $\sigma \in I_q$, $\rho_Q^{\mathrm{univ}}(\sigma)$ is diagonal in this basis. Suppose $\rho_Q^{\mathrm{univ}}(\sigma) = 1_2 + \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $a, b, c, d \in \mathfrak{m}_{R_Q}$. Because ρ_Q^{univ} is tamely ramified at q we see that $\rho_Q^{\mathrm{univ}}(f)\rho_Q^{\mathrm{univ}}(\sigma)\rho_Q^{\mathrm{univ}}(f)^{-1} = \rho_Q^{\mathrm{univ}}(\sigma)^q$. Thus $(\tilde{\alpha}_q/\tilde{\beta}_q - q)b$ and $(\tilde{\beta}_q/\tilde{\alpha}_q - q)c$ lie in $\mathfrak{m}_{R_Q}(b, c)$ so that $(b, c) = \mathfrak{m}_{R_Q}(b, c)$ and (by Nakayama's lemma) $b = c = 0$. □

We let $\xi_{q,Q}$ denote the character $\xi : G_q \to R_Q^\times$ in the conclusion of the lemma. The restriction of $\xi_{q,Q}$ to I_q factors through χ_q. We let ξ_Q denote the character $G_{\mathbb{Q}} \to R_Q^\times$ which is unramified outside the primes of Q and whose restriction to I_q is $\xi_{q,Q}$ for each q in Q. Thus ξ_Q factors through χ_Q. We wish to regard R_Q as an $\mathcal{O}[\Delta_Q]$-algebra, and it will be most convenient to do so via the map which gives rise to ξ_Q^{-2}. Note the following consequence of the lemma 2.44.

Corollary 2.45 *The natural map $R_Q \to R_\emptyset$ gives rise to an isomorphism $R_Q/\mathfrak{a}_Q R_Q \xrightarrow{\sim} R_\emptyset$.*

Lemma 2.46 (a) *If $q \in Q$ then $H^0(\mathbb{F}_q, \mathrm{ad}^0\bar{\rho}) = H^0(\mathbb{F}_q, \mathrm{ad}^0\bar{\rho}(1)) = k$ and $H^1(\mathbb{F}_q, \mathrm{ad}^0\bar{\rho}) = H^1(\mathbb{F}_q, \mathrm{ad}^0\bar{\rho}(1)) = k$.*

(b) *R_Q can be topologically generated as an $\mathcal{O}$-algebra by*

$$\#Q + \dim_k H^1_Q(\mathbb{Q}, \mathrm{ad}^0\bar{\rho}(1))$$

elements.

(c) *If*

$$H^1_\emptyset(\mathbb{Q}, \mathrm{ad}^0\bar{\rho}(1)) \xrightarrow{\sim} \bigoplus_{q \in Q} H^1(\mathbb{F}_q, \mathrm{ad}^0\bar{\rho}(1))$$

then $\#Q = \dim_k H^1_\emptyset(\mathbb{Q}, \mathrm{ad}^0\bar{\rho}(1))$ and R_Q can be topologically generated as an $\mathcal{O}$-algebra by $\#Q$ elements.

Proof: The first part is a direct calculation using the fact that Frob_q acts semi-simply on $\mathrm{ad}^0\bar{\rho}$ with eigenvalues $x, 1, x^{-1}$ for some $x \in \bar{k}\backslash\{0,1\}$. The same is true for $\mathrm{ad}^0\bar{\rho}(1)$ as $q \equiv 1 \bmod \ell$. The second and third parts follow from this and corollary 2.43. □

For the proof of the last theorem of the chapter, we shall need two results on finite groups.

Theorem 2.47 (a) *If H is a finite subgroup of $PGL_2(\mathbb{C})$ then H is isomorphic to one of the following groups: the cyclic group C_n of order n ($n \in \mathbb{Z}_{>0}$), the dihedral group D_{2n} of order $2n$ ($n \in \mathbb{Z}_{>1}$), A_4, S_4 or A_5.*

(b) *If H is a finite subgroup of $PGL_2(\bar{\mathbb{F}}_\ell)$ then one of the following holds:*

- *H is conjugate to a subgroup of the upper triangular matrices;*
- *H is conjugate to $PSL_2(\mathbb{F}_{\ell^r})$ or $PGL_2(\mathbb{F}_{\ell^r})$ for some $r \in \mathbb{Z}_{>0}$;*
- *H is isomorphic to A_4, S_4, A_5 or the dihedral group D_{2r} of order $2r$ for some $r \in \mathbb{Z}_{>1}$ not divisible by ℓ.*

In fact we shall only need part (b) which is due to Dickson [Dic2], secs. 255, 260 (see also [Hu] II.8.27), but we have included part (a) for later reference.

Lemma 2.48 *Let $\mathbb{F}$ be a finite field of odd characteristic ℓ. If $\#\mathbb{F} \neq 5$, then*

$$H^1(SL_2(\mathbb{F}), \mathrm{End}^0(\mathbb{F}^2)) = 0.$$

Proof: This is a special case of results of [CPS], table 4.5 (assuming $\#\mathbb{F} \neq 3$). In fact we shall only need it in the case $\ell = 3$, but the proof in the general case is no more difficult, and we sketch it here for the reader's convenience.

We let B (resp. U) denote the group of upper-triangular (resp. unipotent) matrices in $G = SL_2(\mathbb{F})$. Since ℓ does not divide the index of B in G, the restriction homomorphism

$$H^1(G, \mathrm{End}^0(\mathbb{F}^2)) \to H^1(B, \mathrm{End}^0(\mathbb{F}^2))$$

is injective, so it suffices to prove the latter group vanishes. Since ℓ does not divide the index of U in B, we have

$$H^i(B, M) \cong H^i(U, M)^{B/U}$$

all integers $i \geq 0$ and $\mathbb{F}[B]$-modules M. If $\#\mathbb{F} = 3$, then one checks directly that for $M = \mathrm{End}^0(\mathbb{F}_3^2)$,

$$H^1(U, M) \cong \ker N/(\sigma - 1)M = 0$$

where σ generates U and $N = 1 + \sigma + \sigma^2$ on M. If $\#\mathbb{F} > 5$, then one proceeds by writing

$$(0) = M_0 \subset M_1 \subset M_2 \subset M_3 = \mathrm{End}^0(\mathbb{F}^2)$$

as $\mathbb{F}[B]$-modules where the successive quotients M_i/M_{i-1}, (for $i = 1, 2, 3$) are one-dimensional over $\mathbb{F}$. The calculation is then straightforward using long exact sequences of cohomology, except in the case $\#\mathbb{F} = 9$ where one must also check that the one-dimensional space of classes in $H^1(U, M_3/M_2)$ fixed by B/U maps injectively to $H^2(U, M_2/M_1)$ via the connecting homomorphism. □

Theorem 2.49 *Keep the assumptions of the last section and suppose moreover that if* $L = \mathbb{Q}(\sqrt{(-1)^{(\ell-1)/2}\ell})$ *then* $\bar{\rho}|_{G_L}$ *is absolutely irreducible. Then there exists a non-negative integer* r *such that for any* $n \in \mathbb{Z}_{>0}$ *we can find a finite set of primes* Q_n *with the following properties.*

(a) *If* $q \in Q_n$ *then* $q \equiv 1 \bmod \ell^n$.

(b) *If* $q \in Q_n$ *then* $\bar{\rho}$ *is unramified at* q *and* $\bar{\rho}(\mathrm{Frob}_q)$ *has distinct eigenvalues.*

(c) $\#Q_n = r$.

(d) R_{Q_n} *can be topologically generated by* r *elements as a* $\mathcal{O}$*-algebra.*

Proof: Take $r = \dim_k H^1_{\emptyset}(\mathbb{Q}, \mathrm{ad}^0\bar{\rho}(1))$. It suffices to find a set Q_n with the following properties.

(a) If $q \in Q_n$ then $q \equiv 1 \bmod \ell^n$.

(b) If $q \in Q_n$ then $\bar{\rho}$ is unramified at q and $\bar{\rho}(\mathrm{Frob}_q)$ has distinct eigenvalues.

(c) $H^1_{\emptyset}(\mathbb{Q}, \mathrm{ad}^0\bar{\rho}(1)) \xrightarrow{\sim} \bigoplus_{q\in Q_n} H^1(\mathbb{F}_q, \mathrm{ad}^0\bar{\rho}(1))$.

As each $H^1(\mathbb{F}_q, \mathrm{ad}^0\bar{\rho}(1))$ is one dimensional we may replace the last condition with

$$H^1_{\emptyset}(\mathbb{Q}, \mathrm{ad}^0\bar{\rho}(1)) \hookrightarrow \bigoplus_{q\in Q_n} H^1(\mathbb{F}_q, \mathrm{ad}^0\bar{\rho}(1)).$$

Thus what we need show is that for each non-zero class $[\psi] \in H^1_{\emptyset}(\mathbb{Q}, \mathrm{ad}^0\bar{\rho}(1))$ there is a prime q (depending on $[\psi]$) such that

(a) $q \equiv 1 \bmod \ell^n$,

(b) $\bar{\rho}$ is unramified at q and $\bar{\rho}(\mathrm{Frob}_q)$ has distinct eigenvalues,

(c) $\mathrm{res}_q[\psi] \in H^1(\mathbb{F}_q, \mathrm{ad}^0\bar{\rho}(1))$ is nontrivial.

Using the Chebotarev density theorem we see that it will do to find $\sigma \in G_{\mathbb{Q}}$ such that

(a) $\sigma|_{\mathbb{Q}(\zeta_{\ell^n})} = 1$,

(b) $\mathrm{ad}^0\bar{\rho}(\sigma)$ has an eigenvalue other than 1,

(c) $\psi(\sigma) \notin (\sigma - 1)\mathrm{ad}^0\bar{\rho}(1)$.

For $m \geq 0$, let F_m denote the extension of $\mathbb{Q}(\zeta_{\ell^m})$ cut out by $\mathrm{ad}^0\bar{\rho}$; i.e., the field fixed by the kernel of the representation $\mathrm{ad}^0\bar{\rho}$ restricted to $G_{\mathbb{Q}(\zeta_{\ell^m})}$. We will show that $\psi(G_{F_n})$ is non-trivial. For this it suffices to prove that $H^1(\mathrm{Gal}\,(F_n/\mathbb{Q}), \mathrm{ad}^0\bar{\rho}(1)) = (0)$. Consider the inflation-restriction exact sequence

$$(0) \to H^1(\mathrm{Gal}\,(F_0/\mathbb{Q}), (\mathrm{ad}^0\bar{\rho}(1))^{G_{F_0}}) \to H^1(\mathrm{Gal}\,(F_n/\mathbb{Q}), \mathrm{ad}^0\bar{\rho}(1)) \to H^1(\mathrm{Gal}\,(F_n/F_0), \mathrm{ad}^0\bar{\rho}(1))^{G_{\mathbb{Q}}}.$$

Since F_1/F_0 is an extension of degree prime to ℓ, and since $G_{\mathbb{Q}}$ acts trivially on $\mathrm{Gal}\,(F_n/F_1)$, we have:

$$H^1(\mathrm{Gal}\,(F_n/F_0), \mathrm{ad}^0\bar{\rho}(1))^{G_{\mathbb{Q}}} \simeq \mathrm{Hom}(\mathrm{Gal}\,(F_n/F_1), \mathrm{ad}^0\bar{\rho}(1)^{G_{\mathbb{Q}}}). \tag{2.8.1}$$

Since $\bar{\rho}|_{G_L}$ is absolutely irreducible, the cohomology group in equation (2.8.1) vanishes. On the other hand, G_{F_0} acts trivially on $\mathrm{ad}^0\bar{\rho}$ so the first term vanishes as well unless $\mathrm{Gal}\,(F_0/\mathbb{Q})$ has order divisible by ℓ and has $\mathrm{Gal}\,(\mathbb{Q}(\zeta_\ell)/\mathbb{Q})$ as a quotient. Recall that $\mathrm{Gal}\,(F_0/\mathbb{Q})$ is isomorphic to the projective image of $\bar{\rho}$, so by theorem 2.47 we are reduced to the case $\ell = 3$ and the map

$$\mathrm{Gal}\,(F_0/\mathbb{Q}(\zeta_3)) \to PSL_2(\bar{k})$$

has image conjugate to $PSL_2(\mathbb{F}_{3^r})$ for some $r \geq 1$. It suffices to prove that in this case

$$H^1(\mathrm{Gal}\,(F_0/\mathbb{Q}(\zeta_3)), \mathrm{ad}^0\bar{\rho} \otimes_k \bar{k}) = (0),$$

but this follows directly from lemma 2.48.

Now note that since $\bar{\rho}|_{G_L}$ is absolutely irreducible, so is $\bar{\rho}|_{G_{\mathbb{Q}(\zeta_{\ell^n})}}$. Regarding $\psi(G_{F_n})$ as a $\mathrm{Gal}\,(F_n/\mathbb{Q}(\zeta_{\ell^n}))$-submodule of $\mathrm{ad}^0(\bar{\rho})$, we find that some $g \in \mathrm{Gal}\,(F_n/\mathbb{Q}(\zeta_{\ell^n}))$ of order not dividing ℓ fixes a non-zero element of $\psi(G_{F_n})$. Let σ_0 be a lifting of g to $G_{\mathbb{Q}(\zeta_{\ell^n})}$. We will look for $\sigma = \tau\sigma_0$ with $\tau \in G_{F_n}$. We only need choose τ so that $\psi(\tau\sigma_0) = \psi(\tau) + \psi(\sigma_0) \notin (\sigma_0 - 1)\mathrm{ad}^0\bar{\rho}(1)$. This is possible because $\psi(G_F) \not\subset (g-1)\mathrm{ad}^0\bar{\rho}(1)$. □

3 Modular forms and Galois representations

3.1 From modular forms to Galois representations

We suppose in this section that $f = \sum a_n(f)q^n$ is a newform of weight two and level N_f (see definition 1.21). Let K_f denote the number field in $\mathbb{C}$ generated by the Fourier coefficients $a_n(f)$. Let ψ_f denote the character of f, i.e., the homomorphism $(\mathbb{Z}/N_f\mathbb{Z})^\times \to K_f^\times$ defined by mapping d to the eigenvalue of $\langle d \rangle$ on f.

Recall that a construction of Shimura (section 1.7) associates to f an abelian variety A_f of dimension $[K_f : \mathbb{Q}]$. This abelian variety is a certain quotient of $J_1(N_f)$, and the action of the Hecke algebra on $J_1(N_f)$ provides an embedding

$$K_f \hookrightarrow \mathrm{End}_{\mathbb{Q}}(A_f) \otimes \mathbb{Q}.$$

We saw also that for each prime ℓ the Tate module $\mathcal{T}_\ell(A_f) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$ becomes a free module of rank two over $K_f \otimes \mathbb{Q}_\ell$ (lemma 1.48). The action of the Galois group $G_{\mathbb{Q}}$ on the Tate module commutes with that of K_f, so that a choice of basis for the Tate module provides a representation

$$G_{\mathbb{Q}} \to GL_2(K_f \otimes \mathbb{Q}_\ell). \tag{3.1.1}$$

As $K_f \otimes \mathbb{Q}_\ell$ can be identified with the product of the completions of K_f at its primes over ℓ, we obtain from f certain 2-dimensional ℓ-adic representations of $G_{\mathbb{Q}}$.

ℓ-adic representations: In this discussion, we fix a prime ℓ and a finite extension K of $\mathbb{Q}_\ell$. We let $\mathcal{O}$ denote the ring of integers of K, λ the maximal ideal and k the residue field. We shall consider ℓ-adic representations with coefficients in finite extensions of our fixed field K. We regard K as a subfield of $\bar{\mathbb{Q}}_\ell$ and fix embeddings $\bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_\ell$ and $\bar{\mathbb{Q}} \hookrightarrow \mathbb{C}$. If K' is a finite extension of K with ring of integers $\mathcal{O}'$, then we say that an ℓ-adic representation $G_\ell \to GL_2(K')$ is good (respectively, ordinary, semistable) if it is conjugate over K' to a representation $G_\ell \to GL_2(\mathcal{O}')$ which is good (respectively, ordinary, semistable) in the sense of section 2.4.

Let K'_f denote the K-algebra in $\bar{\mathbb{Q}}_\ell$ generated by the Fourier coefficients of f. Thus K'_f is a finite extension of K, and it contains the completion of K_f at the prime over ℓ determined by our choice of embeddings. We let $\mathcal{O}'_f$ denote the ring of integers of K'_f and write k_f for its residue field. We define

$$\rho_f : G_{\mathbb{Q}} \to GL_2(K'_f)$$

as the pushforward of (3.1.1) by the natural map $K_f \otimes \mathbb{Q}_\ell \to K'_f$. We assume the basis is chosen so that ρ_f factors through $GL_2(\mathcal{O}'_f)$. We also let ψ'_f denote the finite order ℓ-adic character

$$G_{\mathbb{Q}} \twoheadrightarrow \mathrm{Gal}\,(\mathbb{Q}(\zeta_{N_f})/\mathbb{Q}) \to (K'_f)^\times$$

obtained from ψ_f.

The following theorem lists several fundamental properties of the ℓ-adic representations ρ_f obtained from Shimura's construction. The result is a combination of the work of many mathematicians. We discuss some of the proofs and provide references below. In the statement we fix f as above and write simply N, a_n, ρ, ψ, ψ' and K' for N_f, $a_n(f)$, ρ_f, ψ_f, ψ'_f and K'_f respectively.

Theorem 3.1 *The ℓ-adic representation*

$$\rho : G_{\mathbb{Q}} \to GL_2(K')$$

has the following properties.

(a) *If $p \nmid N\ell$ then ρ is unramified at p and $\rho(\mathrm{Frob}_p)$ has characteristic polynomial*

$$X^2 - a_p X + p\psi(p).$$

(b) $\det(\rho)$ *is the product of ψ' with the ℓ-adic cyclotomic character ϵ, and $\rho(c)$ is conjugate to $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.*

(c) *ρ is absolutely irreducible.*

(d) *The conductor $N(\rho)$ is the prime-to-ℓ-part of N.*

(e) *Suppose that $p \neq \ell$ and $p \| N$. Let χ denote the unramified character $G_p \to (K')^\times$ satisfying $\chi(\mathrm{Frob}_p) = a_p$. If p does not divide the conductor of ψ, then $\rho|_{G_p}$ is of the form*

$$\begin{pmatrix} \chi\epsilon & * \\ 0 & \chi \end{pmatrix}.$$

If p divides the conductor of ψ, then $\rho|_{G_p}$ is of the form

$$\chi^{-1}\epsilon\psi'|_{G_p} \oplus \chi.$$

(f) *If $\ell \nmid 2N$, then $\rho|_{G_\ell}$ is good. Moreover $\rho|_{G_\ell}$ is ordinary if and only if a_ℓ is a unit in the ring of integers of K', in which case $\rho_{I_\ell}(\mathrm{Frob}_\ell)$ is the unit root of the polynomial $X^2 - a_\ell X + \ell\psi(\ell)$.*

(g) *If ℓ is odd and $\ell \| N$, but the conductor of ψ is not divisible by ℓ, then $\rho|_{G_\ell}$ is ordinary and $\rho_{I_\ell}(\mathrm{Frob}_\ell) = a_\ell$.*

Proof: Part (a) was established by Shimura ([Shi2], [Shi3]). The key ingredient is the Eichler-Shimura congruence relation, theorem 1.29. Recall that $J_1(N)$ has good reduction at primes p not dividing N. So the action of G_p on $\mathcal{T}_\ell(A_f) \otimes_{\mathbf{Z}_\ell} \mathbb{Q}_\ell$ is unramified and is in fact described by the action of $\mathrm{Frob}_p \in G_{\mathbb{F}_p}$ on the Tate module of the reduction. But this is given by the Frobenius endomorphism F whose characteristic polynomial is computed in corollary 1.41.

The first assertion of (b) follows from (a) on applying the Chebotarev density theorem. The second assertion then follows on noting that $\psi(-1) = 1$.

Part (c) was proved by Ribet (see section 2 of [R3]). Assuming reducibility of the representation, he applies algebraicity results of Lang and Serre to obtain a contradiction to the estimate on the Fourier coefficients stated in theorem 1.24.

Parts (d) and (e) follow from a deep result of Carayol [Ca1], Thm. (A), building on the work of Langlands [Ll1], Deligne and others. In fact, this result and the local Langlands correspondence characterize $\rho|_{G_p}$ in terms of $\psi|_{G_p}$ and the L- and ϵ-factors at p of twists of f. The descriptions in the case of $p \| N$ are based on the analysis of Deligne-Rapoport of the reduction mod p of $J_1(N)$ (see [DR], [Ll1]).

The first assertion of (f) follows from the fact that A_f has good reduction at ℓ if ℓ does not divide N. The second assertion of (f) (respectively, all of (g)) follows from the Eichler-Shimura congruence relation (respectively, the work of Deligne-Rapoport), and general results on ℓ-divisible groups and the reduction of abelian varieties; see thm. 2.2 of [W2], lemma 2.1.5 of [W1], §12 of [Gro] and thms. 2.5 and 2.6 of [Edi]. The restriction to odd ℓ in (f) and (g) is made primarily out of lack of suitable definitions. □

Mod ℓ representations: We maintain the notation used in the discussion of ℓ-adic representations. Define

$$\bar{\rho}_f : G_{\mathbb{Q}} \to GL_2(k_f)$$

to be the *semi-simplification* of the reduction of ρ_f. (See the discussion following proposition 2.6.) Assertions analogous to those in theorem 3.1 hold for $\bar{\rho} = \bar{\rho}_f$, except that

- The representation need not be absolutely irreducible (as in (c)). However if ℓ is odd, one checks using (b) that $\bar{\rho}$ is irreducible if and only if it is absolutely irreducible.

- In (d) one only has divisibility of the prime-to-ℓ part of N_f by $N(\bar{\rho})$ in general.

The various possibilities for $m_p(\bar{\rho})$ to be strictly less than the exponent of p in N (where $p \neq \ell$) were classified independently by Carayol [Ca2] and Livné [Liv]. We record the following consequence of their results (cf. the introduction of [DT1]):

Proposition 3.2 *Suppose that p is a prime such that $p|N_f$, $p \not\equiv 1 \bmod \ell$ and $\bar{\rho}_f$ is unramified at p. Then* $\operatorname{tr}(\bar{\rho}_f(\mathrm{Frob}_p))^2 = (p+1)^2$ *in* k_f.

Artin representations: The theory of Hecke operators and newforms (see section 1.3) extends to modular forms on $\Gamma_1(N)$ of arbitrary weight. The construction of ℓ-adic representations associated to newforms was generalized to weight greater than 1 by Deligne [De] using étale cohomology. There are also Galois representations associated to newforms of weight 1 by Deligne and Serre [DS], but an essential difference is that these are Artin representations.

Theorem 3.3 *If $g = \sum a_n(g)q^n$ is a newform of weight one, level N_g and character ψ_g, then there is an irreducible Artin representation*

$$\rho_g : G_{\mathbb{Q}} \longrightarrow GL_2(\mathbb{C})$$

of conductor N_g with the following property: If $p \not| N_g$, then the characteristic polynomial of $\rho_g(\mathrm{Frob}_p)$ is

$$X^2 - a_p(g)X + \psi_g(p).$$

Remark 3.4 Note that $\det(\rho_g)$ is the character of $G_{\mathbb{Q}}$ corresponding to ψ and $\rho_g(c)$ is conjugate to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ (see theorem 3.1).

Remark 3.5 A basis can be chosen so that the representation ρ_g takes values in $GL_2(K_g)$ (where K_g is the number field generated by the $a_n(g)$). Moreover suppose that K is a finite extension of $\mathbb{Q}_\ell$ in $\bar{\mathbb{Q}}_\ell$ and we have fixed embeddings of $\bar{\mathbb{Q}}$ in $\mathbb{C}$ and $\bar{\mathbb{Q}}_\ell$). If K_g is contained in K, then we can view ρ_g as giving rise to an ℓ-adic representation $G_{\mathbb{Q}} \to GL_2(K)$ and hence a mod ℓ representation $G_{\mathbb{Q}} \to GL_2(k)$.

Remark 3.6 A key idea in the construction of ρ_g is to first construct the mod ℓ representations using those already associated to newforms of higher weight. More precisely, suppose that $K_g \hookrightarrow K$ as in remark 3.5. One can show that for some newform f of weight 2 and level N_f dividing $N\ell$ we have

$$a_p(g) \equiv a_p(f), \qquad \psi_g(p)) \equiv p\psi_f(p)$$

for all $p \not| N\ell$, the congruence being modulo the maximal ideal of the ring of integers of K'_f. Thus $\bar{\rho}_f$ is the semisimplification of the desired mod ℓ representation (with scalars extended to k_f).

3.2 From Galois representations to modular forms

It is conjectured that certain types of two-dimensional representations of $G_{\mathbb{Q}}$ always arise from the constructions described in section 3.1. We now state some of the conjectures and the results known prior to [W3] and [TW].

Artin representations:

Conjecture 3.7 *Let $\rho : G_{\mathbb{Q}} \to GL_2(\mathbb{C})$ be a continuous irreducible representation with $\det(\rho(c)) = -1$. Then ρ is equivalent to ρ_g for some newform g of weight one.*

Recall that ρ_g is the Artin representation associated to g by the Deligne-Serre construction (theorem 3.3).

Remark 3.8 Conjecture 3.7 is equivalent to the statement that the Artin L-functions attached to ρ and to all its twists by one-dimensional characters are entire. (The Artin conjecture predicts that the Artin L-function $L(s, \rho)$ is entire, for an arbitrary irreducible, non-trivial Artin representation $\rho : G_{\mathbb{Q}} \longrightarrow GL_d(\mathbb{C})$.)

A large part of conjecture 3.7 was proved by Langlands in [Ll2], and the results were extended by Tunnell [Tu].

Theorem 3.9 *Let $\rho : G_{\mathbb{Q}} \to GL_2(\mathbb{C})$ be a continuous irreducible representation such that $\rho(G_{\mathbb{Q}})$ is solvable and $\det(\rho(c)) = -1$. Then ρ is equivalent to ρ_g for some newform g of weight one.*

Remark 3.10 Note that by theorem 2.47, part (a), the solvability hypothesis excludes only the case where the projective image of ρ is isomorphic to A_5.

Remark 3.11 If the projective image of ρ is dihedral, then ρ is induced from a character of a quadratic extension of $\mathbb{Q}$. In this case the result can already be deduced from work of Hecke.

Mod ℓ representations: We fix notation as in the discussion of ℓ-adic and mod ℓ representations in section 3.1.

Definition 3.12 We say that a representation $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(k)$ is *modular (of level N)* if for some newform f of weight 2 (and level N), $\bar{\rho}$ is equivalent over k_f to $\bar{\rho}_f$.

By proposition 1.32 the notion is independent of the choices in section 3.1 of embeddings $K \hookrightarrow \bar{\mathbb{Q}}_\ell$, $\bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_\ell$ and $\bar{\mathbb{Q}} \hookrightarrow \mathbb{C}$. Moreover if K' is a finite extension of K with residue field k', then $\bar{\rho}$ is modular if and only if $\bar{\rho} \otimes_k k'$ is modular.

The following was conjectured by Serre [Se7], (3.2.3). (See also [Da1] for further discussion and references.)

Conjecture 3.13 *Let $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(k)$ be a continuous absolutely irreducible representation with $\det(\bar{\rho}(c)) = -1$. Then $\bar{\rho}$ is modular.*

Some cases of Serre's conjecture can be deduced from theorem 3.9.

Theorem 3.14 *Let $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(k)$ be a continuous absolutely irreducible representation with $\det(\bar{\rho}(c)) = -1$. Suppose that one of the following holds:*

(a) $k = \mathbb{F}_3$;

(b) *the projective image of $\bar{\rho}$ is dihedral.*

Then $\bar{\rho}$ is modular.

Sketch of proof: For case (a), we consider the surjection

$$GL_2(\mathbb{Z}[\sqrt{-2}]) \to GL_2(\mathbb{F}_3)$$

defined by reduction mod $(1 + \sqrt{-2})$. One checks that there is a section $s : GL_2(\mathbb{F}_3) \to GL_2(\mathbb{Z}[\sqrt{-2}])$ and applies theorem 3.9 to $s \circ \bar{\rho}$. The resulting representation arises from a weight one newform, and hence its reduction $\bar{\rho}$ is equivalent to $\bar{\rho}_f$ for some f (see remark 3.6).

In case (b), $\bar{\rho}$ is equivalent to a representation of the form $\mathrm{Ind}_{G_F}^{G_{\mathbb{Q}}} \bar{\xi}$ where F is a quadratic extension of $\mathbb{Q}$ and $\bar{\xi}$ is a character $G_F \to k^{\times}$. (We have here enlarged K if necessary.) Let n be the order of $\bar{\xi}$, choose an embedding $\mathbb{Q}(e^{2\pi i/n}) \hookrightarrow K$ and lift $\bar{\xi}$ to a character $\xi : G_F \to \mathbb{Z}[e^{2\pi i/n}]^{\times}$. We may always choose ξ so that the Artin representation $\rho = \mathrm{Ind}_{G_F}^{G_{\mathbb{Q}}} \xi$ is odd, i.e., $\det(\rho(c)) = -1$. (In the case $\ell = 2$ and F real quadratic, we may have to multiply ξ by a suitable quadratic character of G_F.) We then apply theorem 3.9 to ρ and deduce as in case (a) that $\bar{\rho}$ is modular. □

Serre also proposed a refinement ([Se7], (3.2.4)) of the conjecture which predicts that $\bar{\rho}$ is associated to a newform of specified weight, level and character. Through work of Mazur, Ribet [R5], Carayol [Ca2], Gross [Gro] and others, this refinement is now known to be equivalent to conjecture 3.13 if ℓ is odd. (One also needs to impose a mild restriction in the case $\ell = 3$.) See [R6] and [Di1] for statements of the results and further references; here we give a variant which applies to newforms of weight two. Before doing so, we assume ℓ is odd and define an integer $\delta(\bar{\rho})$ as follows:

- $\delta(\bar{\rho}) = 0$ if $\bar{\rho}|_{G_\ell}$ is good;
- $\delta(\bar{\rho}) = 1$ if $\bar{\rho}|_{G_\ell}$ is not good and $\bar{\rho}|_{I_\ell} \otimes_k \bar{k}$ is of the form

$$\begin{pmatrix} \epsilon^a & * \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} \epsilon & * \\ 0 & \epsilon^a \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \psi^a & 0 \\ 0 & \psi^{\ell a} \end{pmatrix}$$

  for some positive integer $a < \ell$. (Recall that ϵ is the cyclotomic character and ψ is the character of I_ℓ defined after lemma 2.22.)

- $\delta(\bar{\rho}) = 2$ otherwise.

Theorem 3.15 *Suppose that ℓ is odd and $\bar{\rho}$ is absolutely irreducible and modular. If $\ell = 3$, then suppose also that $\bar{\rho}|_{G_{\mathbf{Q}(\sqrt{-3})}}$ is absolutely irreducible. Then there exists a newform f of weight two such that*

- *$\bar{\rho}$ is equivalent over k_f to $\bar{\rho}_f$;*
- *$N_f = N(\bar{\rho})\ell^{\delta(\bar{\rho})}$;*
- *the order of ψ_f is not divisible by ℓ.*

Proof: The existence of such an f follows from [Di1] thm. 1.1, thm. 5.1 and lemma 2.1, but with N_f dividing $N(\bar{\rho})\ell^{\delta(\bar{\rho})}$. By lemma 2.7 above, we see that N_f is divisible by $N(\bar{\rho})$. The divisibility of N_f by $\delta(\bar{\rho})$ follows from results in sec. 8 of [Gro] and sec. 2.4 of [Edi] (cf. sec. 4 of [Kh]). □

ℓ-adic representations: We again use the notation of section 3.1. Let $\rho : G_{\mathbf{Q}} \to GL_2(K)$ be an ℓ-adic representation.

Definition 3.16 We say that ρ is *modular* if for some weight 2 newform f, ρ is equivalent over K'_f to ρ_f.

The notion is independent of the choices of embeddings and well-behaved under extension of scalars by proposition 1.32 (cf. definition 3.12).

The following is a special case of a conjecture of Fontaine and Mazur [FM].

Conjecture 3.17 *If $\rho : G_{\mathbf{Q}} \to GL_2(K)$ is an absolutely irreducible ℓ-adic representation and $\rho|_{G_\ell}$ is semistable (in the sense of section 2.4), then ρ is modular.*

(Recall that for us ℓ-adic representations are defined to be unramified at all but finitely many primes. Recall also that if $\rho|_{G_\ell}$ is semistable, then by definition $\det \rho|_{I_\ell}$ is the cyclotomic character ϵ.)

Remark 3.18 Relatively little was known about this conjecture before Wiles' work [W3]. Wiles proves that under suitable hypotheses, the modularity of $\bar{\rho}$ implies that of ρ.

Remark 3.19 The conjecture stated in [FM] is stronger than the one here; in particular, the semistability hypothesis could be replaced with a suitable notion of potential semistability. On the other hand, one expects that if $\rho|_{G_\ell}$ is semistable, then it is equivalent to ρ_f (over K'_f) for some f on $\Gamma_1(N(\rho)) \cap \Gamma_0(\ell)$ (and on $\Gamma_1(N(\rho))$ if $\rho|_{G_\ell}$ is good).

The Shimura-Taniyama conjecture: Conjecture 1.54 can be viewed in the framework of the problem of associating modular forms to Galois representations. Let E be an elliptic curve defined over $\mathbb{Q}$. For each prime ℓ, we let $\rho_{E,\ell}$ denote the ℓ-adic representation $G_{\mathbb{Q}} \to GL_2(\mathbb{Q}_\ell)$ defined by the action of $G_{\mathbb{Q}}$ on the Tate module of E (see section 2.2).

Proposition 3.20 *The following are equivalent:*

(a) *E is modular.*

(b) *$\rho_{E,\ell}$ is modular for all primes ℓ.*

(c) *$\rho_{E,\ell}$ is modular for some prime ℓ.*

Proof: If E is modular, then E is isogenous to A_f for some weight two newform f with $K_f = \mathbb{Q}$ (see section 1.8). It follows that for each prime ℓ, $\rho_{E,\ell}$ is equivalent to the ℓ-adic representation ρ_f. Hence (a) $\Rightarrow$ (b) $\Rightarrow$ (c).

To show (c) $\Rightarrow$ (b), suppose that for some ℓ and some f, the representations $\rho_{E,\ell}$ and ρ_f are equivalent. First observe that for all but finitely primes p, we have

$$\mathrm{tr}\,(\rho_f(\mathrm{Frob}_p)) = \mathrm{tr}\,(\rho_{E,\ell}(\mathrm{Frob}_p)).$$

We deduce from proposition 2.11 and theorem 3.1, part (a) that for all but finitely many primes p

$$a_p(f) = p + 1 - \#\bar{E}_p(\mathbb{F}_p) \in \mathbb{Z}. \tag{3.2.1}$$

Applying proposition 2.6, we find that for each prime ℓ, $\rho_{E,\ell}$ is equivalent to ρ_f and is therefore modular.

We finally show that (b) $\Rightarrow$ (a). The equality (3.2.1) holds for all primes p not dividing N_f, which by theorem 3.1, part (d), is the conductor of E. Since $\det(\rho_f) = \det(\rho_{E,\ell}) = \epsilon$, we see by 3.1, Part (b) that ψ_f is trivial. By theorem 1.27 parts (b) and (d) (or [AL] thm. 3), a_p is in $\{0, \pm 1\}$ for primes p dividing N_f. Thus $K_f = \mathbb{Q}$ and A_f is an elliptic curve. Faltings' isogeny theorem (see [CS], sec. II.5) now tells us that E and A_f are isogenous and we conclude that E is modular. □

Remark 3.21 Note that the equivalence (b) $\Leftrightarrow$ (c) does not require Faltings' isogeny theorem.

Proposition 3.22 *If the Fontaine-Mazur conjecture (conjecture 3.17) holds for some prime ℓ, then the Shimura-Taniyama conjecture (conjecture 1.54) holds. If Serre's conjecture (conjecture 3.13) holds for infinitely many ℓ, then conjecture 1.54 holds.*

Proof: The first assertion is immediate from proposition 3.20 and the irreducibility of $\rho_{E,\ell}$. See [Se7], sec. 4.6 for a proof of the second. (We have implicitly chosen the field K to be $\mathbb{Q}_\ell$ in the statements of conjectures 3.17 and 3.13, but it may be replaced by a finite extension.) □

Remark 3.23 Note that to prove a given elliptic curve E is modular, it suffices to prove that conjecture 3.17 holds for a single ℓ at which E has semistable reduction. Wiles' approach is to show that certain cases of conjecture 3.13 imply cases of conjecture 3.17 and hence cases of conjecture 1.54.

3.3 Hecke algebras

In this section fix the following notation. Let ℓ be an odd prime, let K be a finite extension of $\mathbb{Q}_\ell$, let $\mathcal{O}$ denote the ring of integers of K, let λ denote its maximal ideal and k its residue field. Fix embeddings $K \hookrightarrow \bar{\mathbb{Q}}_\ell$, $\bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_\ell$ and $\bar{\mathbb{Q}} \hookrightarrow \mathbb{C}$. Let $\bar{\rho} : G_{\mathbb{Q}} \to GL_2(k)$ denote a continuous representation with the following properties

(a) $\bar{\rho}$ is irreducible,

(b) $\bar{\rho}$ is modular,

(c) $\det \bar{\rho} = \epsilon$,

(d) $\bar{\rho}|_{G_\ell}$ is semi-stable,

(e) and if $p \neq \ell$ then $\#\bar{\rho}(I_p)|\ell$.

Let us first record the following lemma.

Lemma 3.24 *The representation $\bar{\rho}|_{G_L}$ is absolutely irreducible where $L = \mathbb{Q}(\sqrt{(-1)^{(\ell-1)/2}\ell})$.*

Proof: If it were not then we see that $\ell \nmid \#\bar{\rho}(G_{\mathbb{Q}})$ and so $\bar{\rho}$ is unramified at all $p \neq \ell$. Moreover we can check that $\bar{\rho}|_{I_\ell} \sim \left(\begin{smallmatrix} \epsilon & 0 \\ 0 & 1 \end{smallmatrix}\right)$. If $\ell > 3$ we can use theorem 3.15 to deduce that $\bar{\rho}$ is modular of weight 2 and level 1 and hence obtain a contradiction. If $\ell = 3$ we see that the splitting field of $\bar{\rho}$ is everywhere unramified over $\mathbb{Q}(\sqrt{-3})$ and hence must equal $\mathbb{Q}(\sqrt{-3})$, a contradiction. □

Let Σ denote a finite set of primes. For the application to modularity of semistable elliptic curves, it suffices to consider sets Σ contained in $\Sigma_{\bar{\rho}}$ where $\Sigma_{\bar{\rho}}$ is defined as follows

Definition 3.25 For a representation $\bar{\rho}$ as above, we let $\Sigma_{\bar{\rho}}$ denote the set of primes p satisfying

- $p = \ell$ and $\bar{\rho}|_{G_\ell}$ is good and ordinary; or
- $p \neq \ell$ and $\bar{\rho}$ is unramified at p.

We shall sometimes assume that $\Sigma \subset \Sigma_{\bar{\rho}}$ in order to simplify statements and proofs. Let $\mathcal{N}_\Sigma$ denote the set of newforms f such that

$$\rho_f : G_{\mathbb{Q}} \to GL_2(\mathcal{O}'_f)$$

is equivalent to a lifting of $\bar{\rho} \otimes_k k_f$ of type Σ and $\ell^2 \not| N_f$ (the last condition is presumably not necessary, cf. remark 3.19). From theorem 3.1 and lemma 2.7 one deduces the following description of $\mathcal{N}_\Sigma$.

Lemma 3.26 *The set $\mathcal{N}_\Sigma$ consists of newforms f such that*

- $\bar{\rho}_f \cong \bar{\rho} \otimes_k k_f$,
- ψ_f *is trivial,*
- N_f *divides* $\ell^\delta N(\bar{\rho}) \prod_{p \in \Sigma - \{\ell\}} p^{\dim \bar{\rho}^{I_p}}$, *where* $\delta = 0$ *if* $\bar{\rho}$ *is good and* $\ell \notin \Sigma$ *and* $\delta = 1$ *otherwise.*

As $\bar{\rho}$ is modular it follows from theorem 3.15 that for all Σ, $\mathcal{N}_\Sigma \neq \emptyset$. Set $\tilde{\mathbb{T}}_\Sigma = \prod_{f \in \mathcal{N}_\Sigma} \mathcal{O}'_f$. If p is a prime not in Σ and not dividing $\ell N(\bar{\rho})$, we let T_p denote the element $(a_p(f))_f$ in $\tilde{\mathbb{T}}_\Sigma$. Then define $\mathbb{T}_\Sigma$ to be the $\mathcal{O}$-subalgebra of $\tilde{\mathbb{T}}_\Sigma$ generated by the elements T_p for such primes p. Then $\mathbb{T}_\Sigma$ is a complete noetherian local $\mathcal{O}$-algebra with residue field k. Moreover it is reduced and it is a finitely generated free $\mathcal{O}$-module.

Lemma 3.27 *There is a continuous representation*

$$\rho_\Sigma^{\mathrm{mod}} : G_{\mathbb{Q}} \to GL_2(\mathbb{T}_\Sigma)$$

such that if $p \not| \ell N(\bar{\rho})$ and $p \notin \Sigma$ then $\rho_\Sigma^{\mathrm{mod}}$ is unramified at p and we have $\operatorname{tr} \rho_\Sigma^{\mathrm{mod}}(\mathrm{Frob}_p) = T_p$. *Moreover we have the following.*

(a) $\rho_\Sigma^{\mathrm{mod}}$ *is a lift of $\bar{\rho}$ of type Σ and there is a unique surjection $\phi_\Sigma : R_\Sigma \twoheadrightarrow \mathbb{T}_\Sigma$ such that $\rho_\Sigma^{\mathrm{mod}} \sim \phi_\Sigma \circ \rho_\Sigma^{\mathrm{univ}}$.*

(b) *If $\Sigma' \supset \Sigma$ then there is a unique surjection $\mathbb{T}_{\Sigma'} \twoheadrightarrow \mathbb{T}_\Sigma$ such that $\rho_{\Sigma'}^{\mathrm{mod}}$ pushes forward to $\rho_\Sigma^{\mathrm{mod}}$ and T_p maps to T_p for $p \not| \ell N(\bar{\rho})$ and $p \notin \Sigma'$.*

(c) *If K' is a finite extension of K and $\mathbb{T}'_\Sigma$ is constructed in the same way as $\mathbb{T}_\Sigma$ but with K' replacing K then $\mathbb{T}'_\Sigma \cong \mathbb{T}_\Sigma \otimes_{\mathcal{O}} \mathcal{O}_{K'}$.*

Proof. Consider $\tilde{\rho}_\Sigma^{\mathrm{mod}} = \prod \rho_f : G_{\mathbb{Q}} \to GL_2(\tilde{\mathbb{T}}_\Sigma)$. Choose some complex conjugation, c, and conjugate $\bar{\rho}$ so that $\bar{\rho}(c) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. By conjugating $\tilde{\rho}_\Sigma^{\mathrm{mod}}$ we may assume that modulo every maximal ideal it reduces to $\bar{\rho}$ and that $\tilde{\rho}_\Sigma^{\mathrm{mod}}(c) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Call the corresponding basis $\{e_+, e_-\}$. Note that by continuity and the Chebotarev density theorem $\operatorname{tr} \tilde{\rho}_\Sigma^{\mathrm{mod}}$ is valued in $\mathbb{T}_\Sigma$. Thus for any $g \in G_{\mathbb{Q}}$ the diagonal entries of $\tilde{\rho}_\Sigma^{\mathrm{mod}}(g)$ lie in $\mathbb{T}_\Sigma$, because $\operatorname{tr} \tilde{\rho}_\Sigma^{\mathrm{mod}}(g)$ and $\operatorname{tr} \tilde{\rho}_\Sigma^{\mathrm{mod}}(cg)$ both do.

By irreducibility of $\bar{\rho}$ we can find some $\sigma \in G_{\mathbb{Q}}$ such that $\bar{\rho}(\sigma) = \left(\begin{smallmatrix} * & b \\ * & * \end{smallmatrix}\right)$ where $b \neq 0$. Rescaling e_+ we may assume that $\tilde{\rho}_\Sigma^{\mathrm{mod}}(\sigma) = \left(\begin{smallmatrix} a & 1 \\ c & d \end{smallmatrix}\right)$. Then for

all $g \in G_{\mathbb{Q}}$ the lower left entry of $\tilde{\rho}_{\Sigma}^{\mathrm{mod}}(g)$ lies in $\mathbb{T}_{\Sigma}$ (look at the upper left entry of $\tilde{\rho}_{\Sigma}^{\mathrm{mod}}(\sigma g)$). Again using the irreducibility of $\bar{\rho}$ we can find $\tau \in G_{\mathbb{Q}}$ such that $\tilde{\rho}_{\Sigma}^{\mathrm{mod}}(\tau) = \begin{pmatrix} * & * \\ e & * \end{pmatrix}$ where $e \in \mathbb{T}_{\Sigma}^{\times}$. Looking at the lower right entry of $\tilde{\rho}_{\Sigma}^{\mathrm{mod}}(\tau g)$ we see that for any $g \in G_{\mathbb{Q}}$ the upper right entry of $\tilde{\rho}_{\Sigma}^{\mathrm{mod}}(g)$ lies in $\mathbb{T}_{\Sigma}$. Thus $\tilde{\rho}_{\Sigma}^{\mathrm{mod}}$ is now in fact valued in $GL_2(\mathbb{T}_{\Sigma})$ and will be our candidate for $\rho_{\Sigma}^{\mathrm{mod}}$. We leave the verification of the other properties of $\rho_{\Sigma}^{\mathrm{mod}}$ as an exercise. □

Example 3.28 Let $\bar{\rho} = \bar{\rho}_{f_{57B},3}$ where f_{57B} is the newform of level 57 discussed in the example of section 1.6. As f_{57B} is not congruent modulo 3 to any form of level 19 or 3 we see by theorem 3.15 that $\bar{\rho}$ is ramified at 19 and $\bar{\rho}|_{G_3}$ is not good. On the other hand $\bar{\rho}$ is semi-stable. The facts that $\bar{\rho}(\mathrm{Frob}_2)$ has order 8 (see the table below) and $3|\#\bar{\rho}(I_{19})$ (from the discussion above) imply that $\bar{\rho} : G_{\mathbb{Q}} \twoheadrightarrow GL_2(\mathbb{F}_3)$. The table below also shows that $\mathbb{T}_{\emptyset} \cong \{(x,y) \in \mathbb{Z}_3^2 : x \equiv y \bmod 3\}$. There is a unique continuous representation $\rho_{\emptyset}^{\mathrm{mod}} : G_{\mathbb{Q}} \to GL_2(\mathbb{T}_{\emptyset})$ such that if $p \not| 57$ then $\rho_{\emptyset}^{\mathrm{mod}}$ is unramified at p and $\mathrm{tr}\,\rho_{\emptyset}^{\mathrm{mod}}(\mathrm{Frob}_p) = T_p$. We have an isomorphism $\mathbb{T}_{\emptyset}/3\mathbb{T}_{\emptyset} \xrightarrow{\sim} \mathbb{F}_3[\varepsilon]$ (where $\varepsilon^2 = 0$) given by $(x,y) \mapsto x + \frac{y-x}{3}\varepsilon$. Thus we get a representation $\rho : G_{\mathbb{Q}} \to GL_2(\mathbb{F}_3[\varepsilon])$. The following table lists the traces of Frobenius elements for the first few unramified primes under these representations:

p	2	5	7	11	13	17	23	29
$\bar{\rho}$	1	1	0	0	0	0	1	-1
$\rho_{\emptyset}^{\mathrm{mod}}$	$(1,-2)$	$(-2,1)$	$(0,3)$	$(0,-3)$	$(6,-6)$	$(-6,3)$	$(4,4)$	$(2,-10)$
ρ	$1-\varepsilon$	$1+\varepsilon$	ε	$-\varepsilon$	$-\varepsilon$	0	1	$-1-\varepsilon$

Exercise 3.29 Show that there are three algebra homomorphisms $\mathbb{T}_{\emptyset} \to \mathbb{Z}/9\mathbb{Z}$ and hence show that there are at least three liftings of $\bar{\rho}$ of type $\emptyset$ to $\mathbb{Z}/9\mathbb{Z}$.

Exercise 3.30 What is the image of ρ?

We will need two deeper properties of the Hecke algebras $\mathbb{T}_{\Sigma}$ (theorem 3.31 and theorem 3.36 below). These results will be proved in the next chapter (sections 4.3 and 4.4).

Theorem 3.31 *Let Q be a finite set of primes as in section 2.8. Then $\mathbb{T}_Q$ is a free $\mathcal{O}[\Delta_Q]$-module, where $\mathbb{T}_Q$ inherits the structure of an $\mathcal{O}[\Delta_Q]$-module from R_Q via the homomorphism $\phi_Q : R_Q \twoheadrightarrow \mathbb{T}_Q$.*

Corollary 3.32 $\mathbb{T}_{\emptyset} = \mathbb{T}_Q/\mathfrak{a}_Q$.

Proof: From the definitions and corollary 2.45 we have that

$$\mathbb{T}_{\emptyset} \otimes_{\mathcal{O}} K = (\mathbb{T}_Q \otimes_{\mathcal{O}} K)/\mathfrak{a}_Q,$$

and from the theorem we have that $\mathbb{T}_Q/\mathfrak{a}_Q$ is torsion free. The corollary follows. □

For the second of these results we will need some additional notation and we restrict our attention to sets of primes contained in $\Sigma_{\bar{\rho}}$. Suppose that $\Sigma \subset \Sigma_{\bar{\rho}}$ and that f is an element of $\mathcal{N}_\Sigma$ whose Fourier coefficients are in $\mathcal{O}$. Then $\mathcal{O}'_f = \mathcal{O}$ and projection to the component corresponding to f gives rise to an $\mathcal{O}$-algebra homomorphism

$$\pi = \pi_f : \mathbb{T}_\Sigma \to \mathcal{O}.$$

The pushforward of $\rho_\Sigma^{\mathrm{mod}}$ by π is equivalent to

$$\rho_f : G_{\mathbb{Q}} \to GL_2(\mathcal{O}).$$

Remark 3.33 Most of the objects considered in the rest of the section will depend on the choice of newform f. We also remark that to give an $\mathcal{O}$-algebra homomorphism $\mathbb{T}_\Sigma \to \mathcal{O}$ is equivalent to giving a newform f in $\mathcal{N}_\Sigma$ with coefficients in $\mathcal{O}$. Indeed given such a homomorphism, there exists a newform f in $\mathcal{N}_\Sigma$ such that the homomorphism is defined by $T_p \mapsto a_p(f)$ for all $p \notin \Sigma$ with p not dividing $\ell N(\bar{\rho})$. The fact that $a_p(f) \in \mathcal{O}$ for such p implies (using for example parts (b) and (d) of theorem 1.27 and lemma 4.1 below) that all the Fourier coefficients of f are in $\mathcal{O}$. The uniqueness of f is a consequence of the theory of newforms, theorem 1.22.

For Σ' satisfying

$$\Sigma \subset \Sigma' \subset \Sigma_{\bar{\rho}},$$

let $\pi_{\Sigma'}$ denote the composite $\mathbb{T}_{\Sigma'} \to \mathbb{T}_\Sigma \to \mathcal{O}$. Let

$$\eta_{\Sigma'} = \pi_{\Sigma'}(\mathrm{Ann}_{\mathbb{T}_{\Sigma'}}(\ker \pi_{\Sigma'})). \tag{3.3.1}$$

Note that because $\mathbb{T}_{\Sigma'}$ is reduced, $\eta_{\Sigma'} \neq (0)$. Also let $\wp_{\Sigma'}$ denote the kernel of $\phi_{\Sigma'} \circ \pi_{\Sigma'}$. Recall that

$$\#\wp_{\Sigma'}/\wp_{\Sigma'}^2 = \#H^1_{\Sigma'}(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}).$$

Remark 3.34 We have not yet shown that these groups are finite, but if either is finite, then so is the other and their cardinality is the same.

Note that if ℓ is in $\Sigma_{\bar{\rho}} - \Sigma$ then $T_\ell := (a_\ell(g))_g \in \tilde{\mathbb{T}}_\Sigma$ is actually in $\mathbb{T}_\Sigma$. This follows from theorem 3.1(f), which shows that $T_\ell = \alpha_\ell + \ell\alpha_\ell^{-1}$ where α_ℓ is the eigenvalue of Frob_ℓ on the free rank one $\mathbb{T}_\Sigma$-module $M^{(0)}$ where M is the module underlying $\rho_\Sigma^{\mathrm{mod}}$ (see lemma 2.20). Now that we have shown that $T_\ell \in \mathbb{T}_\Sigma$, we may characterize α_ℓ as the unit root in $\mathbb{T}_\Sigma$ of

$$X^2 - T_\ell X + \ell = 0. \tag{3.3.2}$$

For primes p in $\Sigma_{\bar{\rho}} - \Sigma$ we define an element $c_p \in \mathbb{T}_\Sigma$ by

$$c_p = (p-1)(T_p^2 - (p+1)^2).$$

Note that $\pi(c_p) = (p-1)(a_p(f)^2 - (p+1)^2)$, which is non-zero by theorem 1.27(a). A calculation using theorem 2.17 and theorem 3.1(a) shows that if p is in $\Sigma_{\bar{\rho}} - (\Sigma \cup \ell)$ then

$$\#H^1(G_p, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O})/H^1(G_p/I_p, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) \\ = \#H^0(G_p, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}(1)) = \#\mathcal{O}/\pi(c_p).$$

If ℓ is in $\Sigma_{\bar{\rho}} - \Sigma$ then we see from proposition 2.27 and theorem 3.1(f) that

$$\#H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O})/H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) = \#\mathcal{O}/\pi(c_\ell).$$

For the next proposition define groups H_p for $p \in \Sigma_{\bar{\rho}} - \Sigma$ by

- $H_p = H^1(G_p, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O})/H^1(G_p/I_p, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O})$ if $p \neq \ell$
- and $H_\ell = H^1_{\mathrm{ss}}(G_\ell, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O})/H^1_{\mathrm{f}}(G_\ell, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O})$.

Then we have the following result.

Proposition 3.35 *If $\Sigma \subset \Sigma' \subset \Sigma_{\bar{\rho}}$, then*

$$\#H^1_{\Sigma'}(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O})/H^1_{\Sigma}(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) \leq \#(\mathcal{O}/\pi(\prod_{p \in \Sigma' - \Sigma} c_p)).$$

Moreover if we have equality then the sequence

$$(0) \to H^1_{\Sigma}(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) \to H^1_{\Sigma'}(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) \to \bigoplus_{p \in \Sigma' - \Sigma} H_p \to (0)$$

is exact.

Finally we state the second theorem on Hecke algebras which shall be proved in section 4.4.

Theorem 3.36 *If $\Sigma \subset \Sigma' \subset \Sigma_{\bar{\rho}}$ and f is a newform in $\mathcal{N}_\Sigma$ with coefficients in $\mathcal{O}$, then we have*

$$\eta_{\Sigma'} \subset \pi(\prod_{p \in \Sigma' - \Sigma} c_p)\eta_\Sigma.$$

Corollary 3.37 *With the above notation*

$$\#H^1_{\Sigma'}(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O})/H^1_{\Sigma}(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) \leq \#(\eta_\Sigma/\eta'_\Sigma).$$

3.4 Isomorphism criteria

The main thrust of Wiles' approach is to prove that in many circumstances the map $\phi_\Sigma : R_\Sigma \to \mathbb{T}_\Sigma$ is an isomorphism. For this we will need two criteria from commutative algebra. The first was found by Wiles [W3] (but is presented here in a slightly stronger form due to Lenstra [Len]); the second was developed by Faltings, from the original arguments of [TW]. In both criteria the notion of complete intersection plays a vital part.

Proofs of all the results in this section, together with some background, references and examples, is given in chapter 5.

Definition 3.38 Suppose that A is an object of $\mathcal{C}_\mathcal{O}$ which is finitely generated and free as an $\mathcal{O}$-module. Then we call A a *complete intersection* if and only if for some $r \in \mathbb{Z}_{\geq 0}$ and some $f_1, ..., f_r \in \mathcal{O}[[X_1, ..., X_r]]$ we have

$$A \cong \mathcal{O}[[X_1, ..., X_r]]/(f_1, ..., f_r)$$

(i.e. there are the same number of generators as relations).

We first record a lemma about complete intersections.

Lemma 3.39 *Suppose that K'/K is a finite extension with ring of integers $\mathcal{O}'$ and that A is an object of $\mathcal{C}_\mathcal{O}$ which is finitely generated and free as an $\mathcal{O}$-module. Then A is a complete intersection if and only if $A \otimes_\mathcal{O} \mathcal{O}'$ is.*

For the proof, see chapter 5, lemma 5.30.

Now fix objects R and T of $\mathcal{C}_\mathcal{O}$ and a surjection of $\mathcal{O}$-algebras $\phi : R \twoheadrightarrow T$. Also assume that T *is a finitely generated, free $\mathcal{O}$-module.* The first criterion is as follows.

Theorem 3.40 *Suppose that $\pi : T \twoheadrightarrow \mathcal{O}$. Let $\wp = \ker(\pi \circ \phi) \lhd R$ and let $\eta = \pi(\mathrm{Ann}_T(\ker \pi)) \lhd \mathcal{O}$. Suppose also that $\eta \neq (0)$. Then the following are equivalent.*

(a) *The inequality $\#\wp/\wp^2 \leq \#\mathcal{O}/\eta$ is satisfied.*

(b) *The equality $\#\wp/\wp^2 = \#\mathcal{O}/\eta$ is satisfied.*

(c) *The rings R and T are complete intersections, and the map $\phi : R \longrightarrow T$ is an isomorphism.*

The proof is explained in chapter 5, sections 5.1 to 5.8. (See theorem 5.3.)

For the second criterion let us also fix a non-negative integer r. If $J \lhd \mathcal{O}[[S_1, ..., S_r]]$ is an ideal contained in $(S_1, ..., S_r)$, then by a J-structure we mean a commutative diagram in $\mathcal{C}_\mathcal{O}$

$$\begin{array}{ccccc} & & \mathcal{O}[[S_1, ..., S_r]] & & \\ & & \downarrow & & \\ \mathcal{O}[[X_1, ..., X_r]] & \twoheadrightarrow & R' & \twoheadrightarrow & T' \\ & & \downarrow & & \downarrow \\ & & R & \twoheadrightarrow & T, \end{array}$$

such that

(a) $T'/(S_1, ..., S_r) \xrightarrow{\sim} T$ and $R'/(S_1, ..., S_r) \twoheadrightarrow R$,

(b) for each ideal $I \supset J$, $I = \ker(\mathcal{O}[[S_1, ..., S_r]] \to T'/I)$.

Theorem 3.41 *Suppose there exist a sequence of ideals $J_n \lhd \mathcal{O}[[S_1, ..., S_r]]$ such that $J_0 = (S_1, ..., S_r)$, $J_n \supset J_{n+1}$, $\bigcap_n J_n = (0)$ and for each n there exists a J_n-structure. Then the map $R \longrightarrow T$ is an isomorphism and these rings are complete intersections.*

The proof of this statement is explained in sections 5.9 and 5.10.

3.5 The main theorem

We are now in a position to deduce the main theorems. We will keep the notation and assumptions from the start of section 3.3.

Theorem 3.42 *Keep the notation and assumptions of section 3.3. Then, for all finite sets $\Sigma \subset \Sigma_{\bar{\rho}}$, $\phi_\Sigma : R_\Sigma \longrightarrow \mathbb{T}_\Sigma$ is an isomorphism and these rings are complete intersections.*

Remark 3.43 There seems to be a deep link between the fact that ϕ_Σ is an isomorphism and the fact that $\mathbb{T}_\Sigma$ is a complete intersection. The proof of the theorem divides into two parts. One first proves it in the minimal case where $\Sigma = \emptyset$. One then deduces the full theorem from this special case by a different argument. In both these steps the facts that ϕ_Σ is an isomorphism and that $\mathbb{T}_\Sigma$ is a complete intersection are proved simultaneously.

Proof of theorem 3.42: Note that to prove the theorem we may extend scalars if necessary (by lemma 3.39) and hence assume that both of the following hold:

- The eigenvalues of all elements of the image of ρ are rational over k.
- There is a newform f in $\mathcal{N}_\emptyset$ with coefficients in $\mathcal{O}$, hence an $\mathcal{O}$-algebra homomorphism $\mathbb{T}_\emptyset \to \mathcal{O}$.

We first prove that $\phi_\emptyset : R_\emptyset \longrightarrow \mathbb{T}_\emptyset$ is an isomorphism and that the rings $R_\emptyset$ and $\mathbb{T}_\emptyset$ are complete intersections.

Note that according to theorem 2.49 and lemma 3.24 we can find an integer $r \geq 0$ and for each $n \in \mathbb{Z}_{>0}$ we can find a set Q_n of r primes such that

- if $q \in Q_n$ then $q \equiv 1 \bmod \ell^n$;
- if $q \in Q_n$ then $\bar{\rho}$ is unramified at q and $\bar{\rho}(\mathrm{Frob}_q)$ has distinct eigenvalues;

- R_{Q_n} can be topologically generated by r elements as a $\mathcal{O}$-algebra.

Let $J_n' = ((S_i+1)^{\ell^{n_i}} - 1 : i = 1, \dots, r) \lhd \mathcal{O}[[S_1, \dots, S_r]]$, where the n_i are chosen such that $\mathcal{O}[[S_1, \dots, S_r]]/J_n \cong \mathcal{O}[\Delta_{Q_n}]$. Then for each n we have a diagram

$$\begin{array}{ccccc} & & \mathcal{O}[[S_1, \dots, S_r]] & & \\ & & \downarrow & & \\ \mathcal{O}[[X_1, \dots, X_r]] & \twoheadrightarrow & R_{Q_n} & \twoheadrightarrow & \mathbb{T}_{Q_n} \\ & & \downarrow & & \downarrow \\ & & R_\emptyset & \twoheadrightarrow & \mathbb{T}_\emptyset, \end{array}$$

where

(a) $R_{Q_n}/(S_1, \dots, S_r) \xrightarrow{\sim} R_\emptyset$ (see corollary 2.45),

(b) $\mathbb{T}_{Q_n}$ is finite and free over $\mathcal{O}[[S_1, \dots, S_r]]/J_n'$ (see theorem 3.31),

(c) $\mathbb{T}_{Q_n}/(S_1, \dots, S_r) \xrightarrow{\sim} \mathbb{T}_\emptyset$ (see corollary 3.32).

Let $J_n = ((S_i+1)^{\ell^n} - 1 : i = 1, \dots, r)$. Replacing $\mathbb{T}_{Q_n}$ and R_{Q_n} by $\mathbb{T}_{Q_n}/J_n$ and R_{Q_n}/J_n we see that we have a J_n-structure for every n. Theorem 3.42 for $\Sigma = \emptyset$ now follows from the criterion of theorem 3.41.

We now turn to the proof of theorem 3.42 in the general case. By theorem 2.41 and theorem 3.40 we see that

$$\#H^1_\emptyset(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) = \#\mathcal{O}/\eta_\emptyset, \tag{3.5.1}$$

and so applying corollary 3.37 we see that for any $\Sigma \subset \Sigma_{\bar{p}}$

$$\#H^1_\Sigma(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) \leq \#\mathcal{O}/\eta_\Sigma. \tag{3.5.2}$$

A second application of theorems 2.41 and 3.40 allows us to deduce theorem 3.42. □

Remark 3.44 In certain cases where $\eta_\emptyset = (1)$, the bound (3.5.1) on the order of the Selmer group $H^1_\emptyset(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O})$ also follows from the previous work of Flach, by a different method. See [Fl1] for details.

Corollary 3.45 *Keep the notation of theorem 3.42 and suppose that f is a newform in $\mathcal{N}_\Sigma$ with coefficients in $\mathcal{O}$.*

(a) *We have*

$$\#H^1_\Sigma(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) = \#\mathcal{O}/\eta_\Sigma < \infty,$$

where η_Σ was defined in equation (3.3.1) after remark 3.33.

(b) *If $\Sigma \subset \Sigma' \subset \Sigma_{\bar{p}}$ then*

$$\begin{aligned}(0) \to H^1_\Sigma(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) &\to H^1_{\Sigma'}(\mathbb{Q}, \mathrm{ad}^0\rho_f \otimes_{\mathcal{O}} K/\mathcal{O}) \\ &\to \textstyle\bigoplus_{p \in \Sigma' - \Sigma} H_p \to (0)\end{aligned}$$

is exact, where the groups H_p and H_ℓ were defined before proposition 3.35.

Proof: The first part now follows as a direct consequence of theorem 3.42 and another application of theorems 2.41 and 3.40. The second part follows from the first, together with proposition 3.35 and theorem 3.36. □

Corollary 3.46 *Suppose $\rho : G_{\mathbb{Q}} \to GL_2(K)$ is a continuous representation and let $\bar{\rho}$ denote its reduction. Suppose also that*

(a) *$\bar{\rho}$ is irreducible and modular,*

(b) *if $p \neq \ell$ then $\rho|_{I_p} \sim \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$,*

(c) *$\rho|_{G_\ell}$ is semi-stable,*

(d) *$\det \rho = \epsilon$.*

Then ρ is modular.

Proof: We let Σ denote the set of primes in $\Sigma_{\bar{\rho}}$ at which ρ is ramified. Then $\rho : G_{\mathbb{Q}} \to GL_2(\mathcal{O})$ is a deformation of $\bar{\rho}$ of type Σ, so there is an $\mathcal{O}$-algebra homomorphism $R_\Sigma \to \mathcal{O}$ such that $\rho = \rho_\Sigma^{\text{univ}} \otimes_{R_\Sigma} \mathcal{O}$. Since $\phi_\Sigma : R_\Sigma \to \mathbb{T}_\Sigma$ is an isomorphism by theorem 3.42, it follows that there is a homomorphism $\mathbb{T}_\Sigma \to \mathcal{O}$ sending T_p to $\operatorname{tr}(\rho(\text{Frob}_p))$. Since such a homomorphism is necessarily of the form $T_p \mapsto a_p(f)$ for some newform f, it follows that ρ is equivalent to ρ_f and hence is modular. □

Corollary 3.47 *Suppose that $E/\mathbb{Q}$ is a semi-stable elliptic curve such that $\bar{\rho}_{E,3}$ is irreducible. Then E is modular.*

Proof: One need only apply the last corollary with $\ell = 3$ and theorem 3.14. □

3.6 Applications

The Shimura-Taniyama conjecture for semi-stable elliptic curves:

Theorem 3.48 *If $E/\mathbb{Q}$ is a semistable elliptic curve, then E is modular.*

Proof: By corollary 3.47, it is enough to show that E is modular when its associated mod 3 representation $\bar{\rho}_{E,3}$ is *reducible*, i.e., when E has a subgroup of order 3 defined over $\mathbb{Q}$. Consider the group $E[5]$ of 5-division points of E. The mod 5 Galois representation $\bar{\rho}_{E,5}$ associated to $E[5]$ is irreducible: for otherwise, E would have a subgroup of order 15 defined over $\mathbb{Q}$, and would give rise to a (non-cuspidal) rational point on the modular curve $X_0(15)$. This curve is of genus one, and is known to have only 4 non-cuspidal rational points, which do not correspond to semi-stable elliptic curves (and, at any rate, are known to correspond to modular elliptic curves). Hence we know that $\bar{\rho}_{E,5}$ satisfies all the assumptions of corollary 3.46, except the (crucial!) modularity property. To show that $\bar{\rho}_{E,5}$ is modular, one starts with

Lemma 3.49 *There is an auxiliary (semi-stable) elliptic curve $A/\mathbb{Q}$ which satisfies*

(a) $A[5] \simeq E[5]$ *as $G_{\mathbb{Q}}$-modules.*

(b) $A[3]$ *is an irreducible $G_{\mathbb{Q}}$-module.*

Proof: Let $Y'(5)$ be the curve over $\mathbb{Q}$ which classifies elliptic curves A together with an isomorphism $E[5] \simeq A[5]$ compatible with Weil pairings. Elliptic curves over $\mathbb{Q}$ satisfying (a) correspond to rational points in $Y'(5)(\mathbb{Q})$. Adjoining a finite set of points to $Y'(5)$ yields its compactification $X'(5)$ which is a twist of the modular curve $X(5)$ with full level 5 structure. (I.e., it becomes isomorphic to this curve, over $\bar{\mathbb{Q}}$.) As was shown by Klein, the modular curve $X(5)$ over $\mathbb{C}$ has genus 0. Since $X'(5)$ has a point x_0 defined over $\mathbb{Q}$ corresponding to E, it is isomorphic over $\mathbb{Q}$ to $\mathbb{P}^1$. The rational points of $Y'(5)$ therefore give a plentiful supply of elliptic curves satisfying condition (a). Now consider the curve $Y'(5,3)$ classifying elliptic curves A with an isomorphism $E[5] \simeq A[5]$ (respecting Weil pairings) and a subgroup of A of order 3. One checks that the compactification of $Y'(5,3)$ has genus greater than 1, hence has only finitely many rational points by Faltings' theorem (the Mordell conjecture). It follows that only finitely many points $Y'(5)(\mathbb{Q})$ are in the image of $Y'(5,3)(\mathbb{Q})$ under the natural map $Y'(5,3) \to Y'(5)$. Hence for all but finitely many points x in $Y'(5)(\mathbb{Q})$, the corresponding elliptic curve A satisfies (b) since it has no rational subgroup of order 3. Choosing x arbitrarily close in the 5-adic topology to x_0, we find that the elliptic curve A associated to x is semistable and satisfies the two conditions in the lemma. □

We can now finish the proof of theorem 3.48. Applying corollary 3.47 to the curve A, we find that A is modular. Hence so is the mod 5 representation $\bar{\rho}_{A,5} \simeq \bar{\rho}_{E,5}$. Now applying corollary 3.46 with $\ell = 5$ and ρ the representation of $G_{\mathbb{Q}}$ acting on the 5-adic Tate module of E, we find that $\rho_{E,5}$ is modular, and hence, so is E, as was to be shown.

Remark 3.50 Wiles' original argument uses Hilbert's irreducibility theorem where we have used Faltings' theorem. The alternative presented here is based on a remark of Karl Rubin.

Remark 3.51 The results of [W3] and [TW] actually apply to a larger class of elliptic curves than those which are semistable. In [Di2], their methods are further strengthened to prove that all elliptic curves which have semi-stable reduction at 3 and 5 are modular.

Remark 3.52 Rubin and Silverberg observed that an elliptic curve of the form $y^2 = x(x-a)(x+b)$ has a twist with semi-stable reduction at all odd primes, hence is modular by [Di2]. In fact it is shown in [DK] that their observation together with the general results of [W3] and [TW] already imply modularity.

Fermat's Last Theorem: As was already mentioned in the introduction, the Shimura-Taniyama conjecture for semi-stable elliptic curves (and, more precisely, for the elliptic curves that arise in Frey's construction explained in section 2.2) implies Fermat's Last Theorem.

More precisely, suppose that there is a non-trivial solution to the Fermat equation $x^\ell+y^\ell = z^\ell$, with $\ell > 3$. By theorem 2.15 the Frey curve constructed from this solution (cf. section 2.2) is a semistable elliptic curve $E/\mathbb{Q}$ whose associated mod ℓ representation $\bar{\rho}_{E,\ell}$ is irreducible, unramified outside 2ℓ and is good at ℓ. Serre's conjecture predicts that $\bar{\rho}_{E,\ell}$ arises from a newform of weight 2 and level 2; the "lowering the level" result of Ribet [R5] (cf. theorem 3.15) actually proves this, once we know that E is modular, i.e., $\bar{\rho}_{E,\ell}$ arises from a modular form of weight 2 and *some* level. But this is a contradiction, since there are no modular forms of weight two and level two: such forms would correspond to holomorphic differentials on the modular curve $X_0(2)$ which is of genus 0. This contradiction completes the proof.

Values of L-functions: Also mentioned in the introduction was the relationship between the calculation of the Selmer group (3.5.1) and certain cases of a conjecture of Bloch-Kato [BK], called the Tamagawa number conjecture. It was in this context that partial results were obtained by Flach in [Fl1] (cf. remark 3.44).

If f is a newform of weight 2, then one can associate to f a certain "symmetric square" L-function $L(\mathrm{Symm}^2 f, s)$. We shall recall the definition in section 4.4 and explain how a method of Hida establishes a relationship between $L(\mathrm{Symm}^2 f, 2)$ and $\mathcal{O}/\eta_\Sigma$ in the setting of corollary 3.45. We may therefore regard part (a) of that corollary as a relationship between $L(\mathrm{Symm}^2 f, 2)$ and the size of a Selmer group. While the result is in the spirit of the Tamagawa number conjecture of [BK], we have not verified that the relevant cases of the conjecture can be deduced from it. We shall however state a partial result in the context of semistable elliptic curves. The reader can consult [Fl1] and [Fl2] for a discussion of the relation to the Tamagawa number conjecture.

Suppose that E is a semistable elliptic curve over $\mathbb{Q}$ of conductor N_E and (minimal) discriminant

$$\Delta_E = \prod_{p|N_E} p^{d_p}.$$

The symmetric square L-function associated to E is defined by

$$L(\mathrm{Symm}^2 E, s) = \prod_p L_p(\mathrm{Symm}^2 E, s)$$

where the Euler factors $L_p(\mathrm{Symm}^2 E, s)$ are defined as follows:

- If $p \nmid N_E$, then

$$L_p(\mathrm{Symm}^2 E, s) = ((1-\alpha_p^2 p^{-s})(1-p^{1-s})(1-\beta_p^2 p^{-s}))^{-1}$$

where α_p and β_p are the roots roots of $X^2 - a_p X + p$ with $a_p = p+1-N_p$ as in section 1.1).

- If $p|N_E$, then $L_p(\mathrm{Symm}^2 E, s) = (1-p^{-s})^{-1}$.

Let

$$\Omega_E = \int_{E(\mathbb{C})} \omega_E^{\mathrm{Neron}} \wedge \bar{\omega}_E^{\mathrm{Neron}}$$

where $\omega_E^{\mathrm{Neron}}$ is the Néron differential defined in section 1.1. Since E is modular by theorem 3.48, a method of Shimura (see [Shi4] and the introduction of [St]) establishes the analytic continuation of $L(\mathrm{Symm}^2 E, s)$ to an entire function and shows that $L(\mathrm{Symm}^2 E, 2)$ is a non-zero rational multiple of $i\pi\Omega_E$. We now explain how to deduce the following theorem from Wiles' results and a formula of Hida, corollary 4.21.

Theorem 3.53 *Suppose that E is a semistable elliptic curve and ℓ is a prime such that*

- *$\bar{\rho}_{E,\ell}$ is irreducible, and*
- *ℓ does not divide $2 \prod_{p|N_E} d_p$.*

Then the ℓ-part of

$$\frac{N_E L(\mathrm{Symm}^2 E, 2)}{\pi i \Omega_E}$$

is the order of

$$H^1_{\emptyset}(\mathbb{Q}, \mathrm{ad}^0 \rho_{E,\ell} \otimes_{\mathbf{Z}_\ell} \mathbb{Q}_\ell/\mathbb{Z}_\ell).$$

Sketch of proof: Since E is semistable, it is modular by theorem 3.48. Letting f denote the associated newform, we have $N_f = N_E$ and $L(\mathrm{Symm}^2 E, s) = L(\mathrm{Symm}^2 f, s)$. Furthermore the representations $\rho_{E,\ell}$ and

$$\rho_f : G_{\mathbb{Q}} \to GL_2(\mathbb{Z}_\ell)$$

are equivalent, since they are equivalent over $\mathbb{Q}_\ell$ and $\bar{\rho} = \bar{\rho}_{E,\ell}$ is irreducible. The conditions on E and ℓ ensure that $\bar{\rho}$ and f satisfy the hypotheses of corollary 3.45 with $\Sigma = \emptyset$, so that

$$\#H^1_{\emptyset}(\mathbb{Q}, \mathrm{ad}^0 \rho_{E,\ell} \otimes_{\mathbf{Z}_\ell} \mathbb{Q}_\ell/\mathbb{Z}_\ell) = \#(\mathbb{Z}_\ell/\eta_{\emptyset}).$$

To apply Hida's formula, corollary 4.21, it remains to relate $\det A$ and Ω_E where the matrix $A \in GL_2(\mathbb{C})$ is defined in section 4.4. Using that E is semistable, ℓ is odd and $\bar{\rho}_{E,\ell}$ is irreducible, one obtains a modular parametrization $\pi : X_0(N_E) \to E$ such that

- $\pi^* : H^1(E, \mathbb{Z}_\ell) \to H^1(X_0(N_E), \mathbb{Z}_\ell)$ has torsion-free cokernel;
- the Manin constant for π is not divisible by ℓ ([Maz2], sec. 4(a)).

One can then verify that $\Omega_E^{-1} \det A$ is an ℓ-adic unit and theorem 3.53 follows from corollary 4.21. □

4 Hecke algebras

4.1 Full Hecke algebras

Suppose that K is a finite extension of $\mathbb{Q}_\ell$ for some prime ℓ. Let $\mathcal{O}$ denote its ring of integers and let $k = \mathcal{O}/\lambda$ where λ is the maximal ideal of $\mathcal{O}$. Fix embeddings $K \hookrightarrow \bar{\mathbb{Q}}_\ell$, $\bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_\ell$ and $\bar{\mathbb{Q}} \hookrightarrow \mathbb{C}$. Recall that we defined "Hecke algebras" over $\mathcal{O}$ in two different contexts:

- In section 1.6 as an algebra $\mathbb{T}_{\mathcal{O}}$ generated by the full set of Hecke operators acting on a space of modular forms;
- In section 3.3 as a certain subring $\mathbb{T}_\Sigma$ of a product of fields of Fourier coefficients of newforms giving rise to the same mod ℓ representation.

The first of these provides a concrete geometric description useful for establishing properties of the fine structure of the algebra; the second yields a reduced ring which is more easily interpreted as the coefficient ring of a Galois representation. In the next section we shall relate the two notions by identifying the "reduced Hecke algebras" of the form $\mathbb{T}_\Sigma$ as localizations of the "full Hecke algebras" of the form $\mathbb{T}_{\mathcal{O}}$. Before doing so we need to recall some fundamental properties of the algebras $\mathbb{T}_K$, $\mathbb{T}_{\mathcal{O}}$ and $\mathbb{T}_k$.

Let $\Gamma = \Gamma_H(N)$ for some positive integer N and subgroup H of $(\mathbb{Z}/N\mathbb{Z})^\times$ (see section 1.2). We let $\mathbb{T}_{\mathbf{Z}}$ denote the subring of $\mathrm{End}\,(S_2(\Gamma))$ generated by the operators T_n for all positive integers n and $\langle d \rangle$ for all $d \in (\mathbb{Z}/N\mathbb{Z})^\times$. If R is a ring, then $\mathbb{T}_R$ denotes the R-algebra $\mathbb{T}_{\mathbf{Z}} \otimes R$. Recall that $\mathbb{T}_R$ acts faithfully on $S_2(\Gamma, R)$ and is finitely generated and free as an R-module. (This holds for $R = \mathbb{Z}$, hence for arbitrary R.)

We first record the following lemma:

Lemma 4.1 (a) *$\mathbb{T}_R$ is generated as an R-algebra by either of the following sets of elements:*

- *T_n for all positive integers n.*
- *T_p for all primes p and $\langle d \rangle$ for all d in $(\mathbb{Z}/N\mathbb{Z})^\times$.*

(b) *Suppose that D is a positive integer relatively prime to N. If either D is odd or 2 is invertible in R, then $\mathbb{T}_R$ is generated as an R-algebra by either of the following sets of elements:*

- *T_n for all positive integers n relatively prime to D.*
- *T_p for all primes p not dividing D, and $\langle d \rangle$ for all d in $(\mathbb{Z}/N\mathbb{Z})^\times$.*

For a proof of (a), see [DI], prop. 3.5.1; for (b), see p. 491 of [W3].

The spectrum of $\mathbb{T}_{\mathcal{O}}$: First note that $\mathbb{T}_K$ and $\mathbb{T}_k$ are Artinian, hence have only a finite number of prime ideals, all of which are maximal. Since $\mathbb{T}_{\mathcal{O}}$ is finitely generated and free as an $\mathcal{O}$-module, its maximal (resp. minimal)

prime ideals are those lying over the prime λ (resp. (0)) of $\mathcal{O}$. (This follows from the going-up and going-down theorems, [Mat] thms. 9.4 and 9.5, for example.) It follows that the natural maps

$$\mathbb{T}_{\mathcal{O}} \hookrightarrow \mathbb{T}_{\mathcal{O}} \otimes_{\mathcal{O}} K \cong \mathbb{T}_K; \quad \text{and} \quad \mathbb{T}_{\mathcal{O}} \twoheadrightarrow \mathbb{T}_{\mathcal{O}} \otimes_{\mathcal{O}} k \cong \mathbb{T}_k$$

induce bijections

$$\begin{array}{l} \{\text{maximal ideals of } \mathbb{T}_K\} \leftrightarrow \{\text{minimal primes of } \mathbb{T}_{\mathcal{O}}\} \quad \text{and} \\ \{\text{maximal ideals of } \mathbb{T}_k\} \leftrightarrow \{\text{maximal primes of } \mathbb{T}_{\mathcal{O}}\}. \end{array}$$

Moreover, since $\mathcal{O}$ is complete we have (by [Mat] thms. 8.7 and 8.15, for example) that the natural map

$$\mathbb{T}_{\mathcal{O}} \to \prod_{\mathfrak{m}} \mathbb{T}_{\mathfrak{m}}$$

is an isomorphism, where the product is over the finite set of maximal ideals $\mathfrak{m}$ of $\mathbb{T}_{\mathcal{O}}$ and $\mathbb{T}_{\mathfrak{m}}$ denotes the localization of $\mathbb{T}_{\mathcal{O}}$ at $\mathfrak{m}$. Furthermore each $\mathbb{T}_{\mathfrak{m}}$ is a complete local $\mathcal{O}$-algebra which is finitely generated and free as an $\mathcal{O}$-module, and each minimal prime $\mathcal{P}$ of $\mathbb{T}_{\mathcal{O}}$ is contained in a unique $\mathfrak{m}$.

Now suppose that $f = \sum a_n q^n$ is a normalized eigenform in $S_2(\Gamma, \bar{K})$ for the operators T_n for all $n \geq 1$. Then $T_n \mapsto a_n$ defines a map $\mathbb{T}_{\mathbb{Z}} \to \bar{K}$ and induces a K-algebra homomorphism $\theta_f : \mathbb{T}_K \to \bar{K}$. The image is the finite extension of K generated by the a_n, and the kernel is a maximal ideal of $\mathbb{T}_K$ which depends only on the G_K-conjugacy class of f. Similarly a G_k-conjugacy class of normalized eigenforms in $S_2(\Gamma, \bar{k})$ gives rise to a maximal ideal of $\mathbb{T}_k$.

Recall also that a normalized eigenform f in $S_2(\Gamma, \bar{K})$ has coefficients in $\mathcal{O}_{\bar{K}}$, hence gives rise by reduction to a normalized eigenform $\bar{f}$ in $S_2(\Gamma, \bar{k})$. Furthermore if f and g are G_K-conjugate, then $\bar{f}$ and $\bar{g}$ are G_k-conjugate.

We have thus constructed a diagram of maps of finite sets whose commutativity is easily verified.

$$\begin{array}{ccc} \left\{\begin{array}{c}\text{normalized eigenforms in} \\ S_2(\Gamma, \bar{K}) \text{ modulo } G_K\text{-conjugacy}\end{array}\right\} & \to & \left\{\begin{array}{c}\text{normalized eigenforms in} \\ S_2(\Gamma, \bar{k}) \text{ modulo } G_k\text{-conjugacy}\end{array}\right\} \\ \downarrow & & \downarrow \\ \{\text{maximal ideals of } \mathbb{T}_K\} & & \{\text{maximal ideals of } \mathbb{T}_k\} \\ \updownarrow & & \updownarrow \\ \{\text{minimal primes of } \mathbb{T}_{\mathcal{O}}\} & \twoheadrightarrow & \{\text{maximal primes of } \mathbb{T}_{\mathcal{O}}\}. \end{array} \tag{4.1.1}$$

Proposition 4.2 *The vertical maps are bijective, and the horizontal maps are surjective.*

Proof: For the injectivity of the upper-left vertical arrow, note that if $\mathfrak{p}$ is a maximal ideal of $\mathbb{T}_K$, then all K-algebra homomorphisms $\mathbb{T}_K/\mathfrak{p} \hookrightarrow \bar{K}$ are obtained from a single one by composing with an element of G_K. For the surjectivity, let $K' = \mathbb{T}_K/\mathfrak{p}$ and $\mathfrak{p}'$ denote the kernel of the natural K'-algebra homomorphism $\mathbb{T}_{K'} \twoheadrightarrow K'$. Since $\mathbb{T}_{K'}$ acts faithfully on $S_2(\Gamma, K')$,

the localization $S_2(\Gamma, K')_{\mathfrak{p}}$ is non-zero, hence so is $S_2(\Gamma, K')[\mathfrak{p}']$. (For an R-module M and an ideal I of R, we write $M[I]$ for the intersection over the elements r in I of the kernels of $r : M \to M$.) It follows that there is a normalized eigenform f in $S_2(\Gamma, K')$ so that $\mathfrak{p}'$ is the kernel of the θ'_f : $\mathbb{T}_{K'} \to K'$, and therefore $\mathfrak{p}$ is the kernel of θ_f. To prove that the upper-right vertical arrow is bijective, note that the above arguments carry over with K replaced by k. □

Remark 4.3 The surjectivity of the top arrow is called the Deligne-Serre lifting lemma ([DS], lemma 6.11).

Suppose that $\mathfrak{m}$ is a maximal ideal of $\mathbb{T}_{\mathcal{O}}$. Note that the maximal ideals of $\mathbb{T}_K$ mapping to $\mathfrak{m}$ are precisely those $\mathfrak{p}$ for which $\mathfrak{p} \cap \mathbb{T}_{\mathcal{O}}$ is contained in $\mathfrak{m}$. Note also that the natural map

$$\alpha_{\mathfrak{m}} : \mathbb{T}_{\mathfrak{m}} \otimes_{\mathcal{O}} K \to \prod_{\mathfrak{p}} \mathbb{T}_{\mathfrak{p}}$$

is an isomorphism, where the product is over such $\mathfrak{p}$.

It is straightforward to check that the above constructions are well-behaved with respect to replacing the field K by an extension K'. More precisely, for each set $\mathcal{S}$ in the above diagram, there is a natural surjective map from $\varpi : \mathcal{S}' \to \mathcal{S}$ where $\mathcal{S}'$ is defined by replacing K with K', and these maps are compatible with the maps in the diagram. Furthermore the maps

$$\mathbb{T}_{\mathfrak{p}} \otimes_K K' \to \prod_{\mathfrak{p}' \in \varpi^{-1}(\mathfrak{p})} \mathbb{T}_{\mathfrak{p}'} \quad \text{and} \quad \mathbb{T}_{\mathfrak{m}} \otimes_{\mathcal{O}} \mathcal{O}' \to \prod_{\mathfrak{m}' \in \varpi^{-1}(\mathfrak{m})} \mathbb{T}_{\mathfrak{m}'}$$

are isomorphisms by which $\alpha_{\mathfrak{m}} \otimes_K K'$ can be identified with $\prod_{\mathfrak{m}' \in \varpi^{-1}(\mathfrak{m})} \alpha_{\mathfrak{m}'}$.

Associated Galois representations: Suppose that $\mathfrak{p}$ is a maximal ideal of $\mathbb{T}_K$ and $\mathfrak{m}$ is the associated maximal ideal of $\mathbb{T}_{\mathcal{O}}$. By lemma 1.39,

$$T_\ell(J_H(N)) \otimes_{\mathbb{Z}_\ell} K$$

is free of rank two over $\mathbb{T}_K$, so reduction mod $\mathfrak{p}$ yields a two-dimensional vector space over the field $\mathbb{T}_K/\mathfrak{p}$ endowed with an action of $G_{\mathbb{Q}}$. The resulting Galois representation

$$\rho_{\mathfrak{p}} : G_{\mathbb{Q}} \to GL_2(\mathbb{T}_K/\mathfrak{p})$$

is unramified at all primes p not dividing $N\ell$, and for such p the characteristic polynomial of $\rho_{\mathfrak{p}}(\mathrm{Frob}_p)$ is

$$X^2 - T_p X + p\langle p \rangle \mod \mathfrak{p}.$$

If ℓ is odd, then $\bar{\rho}_{\mathfrak{p}}$ is defined over $\mathbb{T}_{\mathcal{O}}/\mathfrak{m}$ and we write $\rho_{\mathfrak{m}}$ for its semisimplification

$$\rho_{\mathfrak{m}} : G_{\mathbb{Q}} \to GL_2(\mathbb{T}_{\mathcal{O}}/\mathfrak{m}).$$

Thus $\rho_{\mathfrak{m}}$ is unramified at primes p not dividing $N\ell$ and the characteristic polynomial of Frob_p is

$$X^2 - T_p X + p\langle p\rangle \mod \mathfrak{m}.$$

Suppose that g is in the G_K-conjugacy class of eigenforms in $S_2(\Gamma)$ corresponding to $\mathfrak{p}$. If g is a newform then $\mathbb{T}_K/\mathfrak{p}$ is isomorphic to the field denoted K'_g in section 3.1 and $\rho_{\mathfrak{p}}$ can be identified with the representation

$$\rho_g : G_{\mathbb{Q}} \to GL_2(K'_g)$$

considered in theorem 3.1. If also ℓ is odd then $\bar{\rho}_g$ is obtained from $\rho_{\mathfrak{m}}$ by extending scalars. More generally suppose that g is not necessarily a newform and consider the associated newform f. Let D denote the product of the primes which divide N but not N_f. Let $\mathbb{T}_K^{(D)}$ denote the K-subalgebra of $\mathbb{T}_K$ generated by the operators T_n for n relatively prime to D and $\langle d\rangle$ for d in $(\mathbb{Z}/N\mathbb{Z})^\times$. Let $\Gamma' = \Gamma_{H'}(N_f)$ where H' is the image of H in $(\mathbb{Z}/N_f\mathbb{Z})^\times$, and let $\mathbb{T}'_K = \mathbb{T}'_{\mathbb{Z}} \otimes K$ where $\mathbb{T}'_{\mathbb{Z}}$ is the Hecke algebra acting on $S_2(\Gamma')$. Restriction of operators defines a natural homomorphism $\mathbb{T}_K^{(D)} \to \mathbb{T}'_K$ which is surjective by lemma 4.1. The composite with $\mathbb{T}'_K \twoheadrightarrow K'_f$ factors through the field $\mathbb{T}_K^{(D)}/(\mathfrak{p} \cap \mathbb{T}_K^{(D)})$, so we may identify K'_f with a subfield of $\mathbb{T}_K/\mathfrak{p}$ and $\rho_{\mathfrak{p}}$ is then equivalent to the extension of scalars of ρ_f. If ℓ is odd, then $\rho_{\mathfrak{m}}$ and $\bar{\rho}_f$ are defined and equivalent over a common subfield of $\mathbb{T}_{\mathcal{O}}/\mathfrak{m}$ and k_f.

The structure of $\mathbb{T}_K$: We now give an explicit description of $\mathbb{T}_K$ in the case that K contains the coefficients of all eigenforms of level dividing N. Let $\mathcal{N}_\Gamma$ denote the set of newforms in $S_2(\Gamma)$, i.e., the set of newforms f of level N_f dividing N such that H is contained in the kernel of the character

$$(\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/N_f\mathbb{Z})^\times \xrightarrow{\psi_f} K^\times.$$

By theorem 1.22

$$S_2(\Gamma, K) = \bigoplus_{f \in \mathcal{N}_\Gamma} S_{K,f},$$

where $S_{K,f}$ is spanned by the linear independent elements

$$\{\, f(a\tau) \mid a \text{ divides } N/N_f \,\}.$$

For each $f = \sum a_n(f)q^n$ in $\mathcal{N}_\Gamma$, let

- $\mathbb{T}_{K,f}$ denote the image of $\mathbb{T}_K$ in $\mathrm{End}_K S_{K,f}$;
- $A_{K,f}$ denote the polynomial ring over K in the variables $u_{f,p}$ indexed by the prime divisors of N/N_f;
- $I_{K,f}$ denote the ideal in $A_{K,f}$ generated by the polynomials

$$P_{f,p}(u_{f,p}) = u_{f,p}^{v_p(N/N_f)-1}(u_{f,p}^2 - a_p(f)u_{f,p} + \psi_f(p)p),$$

for primes p dividing N/N_f (setting $\psi_f(p) = 0$ if p divides N_f).

Consider the K-algebra homomorphism $A_{K,f} \twoheadrightarrow \mathbb{T}_{K,f}$ defined by mapping $u_{f,p}$ to the operator T_p. Since $P_{f,p}$ is the characteristic polynomial of T_p on the span of $\{ f(ap^i\tau) \mid i = 1, \ldots, v_p(N/N_f) \}$ for each a dividing $N/(N_f p^{v_p(N/N_f)})$, we see that $I_{K,f}$ is contained in the kernel of $A_{K,f} \twoheadrightarrow \mathbb{T}_{K,f}$. Taking the product over f in $\mathcal{N}_\Gamma$, we have a surjective K-algebra homomorphism

$$\prod_f A_{K,f}/I_{K,f} \twoheadrightarrow \prod_f \mathbb{T}_{K,f}.$$

Since the natural map $\mathbb{T}_K \to \prod_f \mathbb{T}_{K,f}$ is injective and

$$\dim_K \mathbb{T}_K = \dim_{\mathbb{C}} S_2(\Gamma, \mathbb{C}) = \sum_f \sigma_0(N/N_f) = \sum_f \dim_K A_{K,f}/I_{K,f},$$

we conclude

Lemma 4.4 *There is an isomorphism of K-algebras:*

$$\alpha : \mathbb{T}_K \to \prod_{f \in \mathcal{N}_\Gamma} A_{K,f}/I_{K,f}$$

defined by

- $\alpha(T_p)_f = a_p(f)$ *if p is a prime not dividing N/N_f;*
- $\alpha(T_p)_f = u_{f,p} \bmod I_{K,f}$ *if p is a prime dividing N/N_f;*
- $\alpha(\langle d \rangle)_f = \psi_f(d)$ *if d is relatively prime to N.*

Remark 4.5 It follows that the algebra $\mathbb{T}_K$ is a "complete intersection" over K in the sense that it is a finite-dimensional K-algebra of the form

$$K[X_1, \ldots, X_r]/(P_1, \ldots, P_r)$$

for some r.

4.2 Reduced Hecke algebras

As in section 4.1, we suppose K is a finite extension of $\mathbb{Q}_\ell$ with ring of integers $\mathcal{O}$ and residue field k, and fix embeddings $K \hookrightarrow \bar{\mathbb{Q}}_\ell$, $\bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_\ell$ and $\bar{\mathbb{Q}} \hookrightarrow \mathbb{C}$. We assume in this section that ℓ is odd and we fix a representation

$$\bar{\rho} : G_{\mathbb{Q}} \to GL_2(k)$$

which is modular (definition 3.12). Thus $\bar{\rho}$ is equivalent to $\rho_{\mathfrak{m}}$ (over $\mathbb{T}_{\mathcal{O}}/\mathfrak{m}$) for some $\Gamma = \Gamma_H(N)$ and maximal ideal $\mathfrak{m}$ of $\mathbb{T}_{\mathcal{O}}$.

We suppose also that $\bar{\rho}$ has the properties listed at the beginning of section 3.3 and that Σ is a finite set of primes contained in $\Sigma_{\bar{\rho}}$ (definition 3.25).

We shall show that Γ and $\mathfrak{m}$ can be chosen so that the reduced Hecke algebra $\mathbb{T}_\Sigma$ can be identified with a localization $\mathbb{T}_\mathfrak{m}$ of the full Hecke algebra $\mathbb{T}_\mathcal{O}$. The main result is due to Wiles ([W3], prop. 2.15), but we also explain an important variant ([TW], lemma 1) which arises when considering special sets of primes Q as in section 2.8 above.

$\mathbb{T}_\Sigma$ and the full Hecke algebra: Let $\Gamma = \Gamma_0(N_\Sigma)$ where

$$N_\Sigma = \ell^\delta N(\bar\rho) \prod_{p\in\Sigma-\{\ell\}} p^{\dim \bar\rho^{I_p}} = \ell^\delta \prod_{p|N(\bar\rho)} p \prod_{p\in\Sigma-\{\ell\}} p^2 \tag{4.2.1}$$

with $\delta = 0$ if $\bar\rho$ is good and $\ell \notin \Sigma$ and $\delta = 1$ otherwise.

Suppose that f is a newform in $\mathcal{N}_\Sigma$. Recall that $\mathcal{N}_\Sigma$, defined in section 3.3, is the set of newforms f in $S_2(\Gamma)$ such that $\bar\rho$ is equivalent to $\bar\rho_f$ over k_f (lemma 3.26). Note that these representations are equivalent if and only if

$$a_p(f) \bmod \lambda' = \operatorname{tr}(\bar\rho(\mathrm{Frob}_p)) \qquad \text{for all } p \nmid N_\Sigma \ell,$$

where λ' is maximal ideal of the ring of integers of K'_f. There is then a normalized eigenform g in $S_2(\Gamma, K'_f)$ characterized by

- $a_p(g) = a_p(f)$ if p does not divide N_Σ/N_f;
- $a_p(g) = 0$ if $p \neq \ell$ and p divides N_Σ/N_f;
- $a_\ell(g)$ is the unit root of $X^2 - a_\ell(f)X + \ell$ if ℓ divides N_Σ/N_f.

Note that if ℓ divides N_Σ/N_f, then $\delta = 1$ and $\bar\rho$ is flat, hence $\bar\rho$ is ordinary and $a_\ell(f)$ is a unit by theorem 3.1 (f).

Lemma 4.6 *For f and g as above, the reduction $\bar g$ is the normalized eigenform in $S_2(\Gamma, k')$ characterized by:*

- $a_p(\bar g) = \operatorname{tr} \bar\rho_{I_p}(\mathrm{Frob}_p)$ *if $p = \ell$ or p is not in Σ (recall that $\bar\rho_{I_p}$ denotes the representation on I_p-coinvariants);*
- $a_p(\bar g) = 0$ *otherwise.*

Furthermore, if K is as in lemma 4.4, then g is the unique normalized eigenform in $S_{K,f}$ with this reduction.

Proof: If p is not in $\Sigma \cup \{\ell\}$ then the formula for $a_p(\bar g) = a_p(\bar f)$ follows from theorem 3.1, parts (a) and (e). If $p = \ell$, then we use parts (f) and (g) of theorem 3.1. Otherwise, we have $a_p(\bar g) = 0$ by theorem 1.27(d) (if $p^2|N_f$ and f has trivial character, then $a_p(f) = 0$). The uniqueness follows from the fact that if $p||N_f$, then $a_p(f) = \pm 1$ (by theorem 1.27(b)). □

Note in particular that $\bar g$ has coefficients in k and is independent of the choice of f in $\mathcal{N}_\Sigma$. We let $\mathfrak{m}$ denote the corresponding maximal ideal of $\mathbb{T}_\mathcal{O}$.

Proposition 4.7 *There is an isomorphism* $\mathbb{T}_\Sigma \xrightarrow{\sim} \mathbb{T}_\mathfrak{m}$ *of* $\mathcal{O}$*-algebras such that* $T_p \mapsto T_p$ *for all primes* p *not dividing* $N_\Sigma \ell$.

Proof: Replacing K by a larger field K', we have $\mathbb{T}_\Sigma \otimes_\mathcal{O} \mathcal{O}' \xrightarrow{\sim} \mathbb{T}'_\Sigma$ (lemma 3.27) and $\mathbb{T}_\mathfrak{m} \otimes_\mathcal{O} \mathcal{O}' \xrightarrow{\sim} \mathbb{T}'_{\mathfrak{m}'}$ (since there is a unique maximal ideal $\mathfrak{m}'$ of $\mathbb{T}_{\mathcal{O}'}$ over $\mathfrak{m}$). We are therefore reduced to the case where K is as in lemma 4.4. Note that $\mathcal{N}_\Sigma$ is the set of newforms f in $\mathcal{N}_\Gamma$ such that

$$a_p(\bar{f}) = \mathrm{tr}\,(\bar{\rho}(\mathrm{Frob}_p)) \qquad \text{for all } p \nmid N_\Sigma \ell.$$

(We have $K = K'_f$ and write $\bar{f}$ for f mod λ.)

We now define an isomorphism of K-algebras

$$\kappa : \mathbb{T}_\mathfrak{m} \otimes K \xrightarrow{\sim} \prod_{f \in \mathcal{N}_\Sigma} K,$$

such that for each $f \in \mathcal{N}_\Sigma$:

- $\kappa(T_p)_f = a_p(f)$ if p is not in Σ;
- $\kappa(T_p)_f = 0$ if p is in $\Sigma - \{\ell\}$;
- $\kappa(T_\ell)_f$ is the unit root of $X^2 - a_\ell(f)X + \psi_f(\ell)\ell$ if ℓ is in Σ.

Recall that $\mathbb{T}_\mathfrak{m} \otimes K \cong \prod_\mathfrak{p} \mathbb{T}_\mathfrak{p}$ where $\mathfrak{p}$ runs over the primes of $\mathbb{T}_K$ whose preimage in $\mathbb{T}_\mathcal{O}$ is contained in $\mathfrak{m}$. Thus according to lemma 4.4 we have

$$\mathbb{T}_\mathfrak{m} \otimes K \xrightarrow{\sim} \prod_{f \in \mathcal{N}_\Gamma} \prod_{\mathfrak{p} \in \mathcal{M}_f} (A_{K,f}/I_{K,f})_\mathfrak{p},$$

where $\mathcal{M}_f$ is the set of prime ideals in $A_{K,f}/I_{K,f}$ whose preimage in $\mathbb{T}_\mathcal{O}$ is contained in $\mathfrak{m}$. If f is not in $\mathcal{N}_\Sigma$, then $\mathcal{M}_f$ is empty. If f is in $\mathcal{N}_\Sigma$, then $\mathcal{M}_f$ consists only of the kernel $\mathfrak{p}_f$ of the map defined by $u_{p,f} \mapsto a_p(g)$ where g is the eigenform of lemma 4.6. Furthermore the maps $u_{p,f} \mapsto a_p(g)$ induce isomorphisms $(A_{K,f}/I_{K,f})_{\mathfrak{p}_f} \xrightarrow{\sim} K$, and taking the product over $f \in \mathcal{N}_\Sigma$, we obtain the desired isomorphism κ.

Identifying $\mathbb{T}_\Sigma$ with the $\mathcal{O}$-subalgebra of $\prod_{f \in \mathcal{N}_\Sigma} K$ generated by the elements $T_p = (a_p(f))_f$ for all p not dividing $N_\Sigma \ell$, it suffices to prove that $\mathbb{T}_\Sigma$ contains $\kappa(T_p)$ for all p dividing $N(\bar{\rho})\ell$. Now observe that for each f in $\mathcal{N}_\Sigma$ the representation ρ_f is isomorphic to

$$G_\mathbb{Q} \to GL_2(\mathbb{T}_\Sigma) \to GL_2(K)$$

obtained by composing $\rho_\Sigma^{\mathrm{mod}}$ with the projection $\mathbb{T}_\Sigma \to K$. It follows that for p dividing $N(\bar{\rho})\ell^\delta$, the image of $(\rho_\Sigma^{\mathrm{mod}})_{I_p}(\mathrm{Frob}_p)$ in K is $(\rho_f)_{I_p}(\mathrm{Frob}_p) = a_p(f)_f$ and therefore

$$(\rho_\Sigma^{\mathrm{mod}})_{I_p}(\mathrm{Frob}_p) = \kappa(T_p).$$

Finally if $\delta = 0$, then we conclude from lemma 4.1 that $\mathbb{T}_\Sigma$ contains $\kappa(T_\ell)$. □

A twisted variant: Suppose now that Σ is a finite set of primes contained in $\Sigma_{\bar\rho} - \{\ell\}$ such that if p is in Σ then $\bar\rho(\mathrm{Frob}_p)$ has eigenvalues α_p and β_p in k satisfying

$$\alpha_p/\beta_p \neq p^{\pm 1}. \tag{4.2.2}$$

Note that (4.2.2) is equivalent to

$$\mathrm{tr}\,(\bar\rho(\mathrm{Frob}_p))^2 \neq (p+1)^2. \tag{4.2.3}$$

Let Q denote the set of primes in Σ such that $q \equiv 1 \bmod \ell$. The set of primes Q is therefore as in section 2.8. Choosing an eigenvalue α_q of $\bar\rho(\mathrm{Frob}_q)$ for each $q \in Q$ as in section 2.8, we regard R_Q and hence $\mathbb{T}_Q$ as an $\mathcal{O}[\Delta_Q]$-algebra. (Recall that Δ_Q is the the maximal quotient of $(\mathbb{Z}/(\prod_{q\in Q} q)\mathbb{Z})^\times$ of ℓ-power order.)

Instead of working with $\Gamma_0(N_\Sigma)$ as in proposition 4.7, we shall now work with the group

$$\Gamma = \Gamma_0(N_\emptyset) \cap \Gamma_1(M), \tag{4.2.4}$$

where

$$M = \prod_{p\in\Sigma-Q} p^2 \prod_{q\in Q} q.$$

Remark 4.8 We are about to relate $\mathbb{T}_Q$ to a localization of $\mathbb{T}_\mathcal{O}$, where $\mathbb{T}_\mathcal{O}$ is now defined using the Hecke operators on $S_2(\Gamma)$, with Γ defined by (4.2.4). Recall that $\mathbb{T}_Q$ is defined using modular forms with trivial character, but note that the modular forms involved in the definition of $\mathbb{T}_\mathcal{O}$ may have non-trivial character. The purpose of establishing this relationship is to give a concrete realization of the image of Δ_Q in $\mathbb{T}_Q$ for the purpose of proving theorem 3.31.

Suppose that f is a newform pair in $\mathcal{N}_\emptyset$. For each q in Q, let $\tilde\alpha_q(f)$ be the root of $X^2 - a_q(f)X + q = 0$ (in K'_f) whose image in k_f is α_q. We let g denote the unique normalized eigenform in $S_2(\Gamma, K'_f)$ of trivial character such that

- $a_p(g) = a_p(f)$ if p is not in Σ;
- $a_p(g) = \tilde\alpha_p(f)$ if p is in Q;
- $a_p(g) = 0$ otherwise.

The reduction $\bar g$ is the unique normalized eigenform in $S_2(\Gamma, k')$ of trivial character such that

- $a_p(\bar g) = \mathrm{tr}\,\bar\rho_{I_p}(\mathrm{Frob}_p)$ if p is not in Σ;

- $a_p(\bar{g}) = \alpha_p$ if p is in Q;
- $a_p(\bar{g}) = 0$ otherwise.

Thus $\bar{g}$ has coefficients in k and is independent of the choice of f in $\mathcal{N}_\emptyset$. We let $\mathfrak{m}$ denote the corresponding maximal ideal of $\mathbb{T}_{\mathcal{O}}$.

Suppose for the moment that K contains the coefficients of all eigenforms of level dividing N_Σ. Let $\mathcal{N}'$ denote the set of newforms g in $\mathcal{N}_\Gamma$ such that

- $\bar{\rho}_g \sim \bar{\rho}$;
- $a_p(\bar{g}) = \alpha_p$ for all p dividing $N_g/N_\emptyset$.

Suppose we are given a newform $g \in \mathcal{N}'$. Let ψ_g denote its character and write Q_g for the conductor of ψ_g. Note that ψ_g has trivial reduction and hence ℓ-power order, and that Q_g divides Q. By proposition 3.2 we see that if $p \in \Sigma - Q$, then ρ_g is unramified at p and hence N_g is not divisible by p. Furthermore, by theorem 3.1 (e) and (4.2.2), $N_g = N_\emptyset Q_g$. Let ξ_g denote the character of $(\mathbb{Z}/Q_g\mathbb{Z})^\times$ of ℓ-power order such that ξ_g^{-2} is the primitive character associated to ψ_g. Then

$$g \otimes \xi_g = \sum \xi_g(n) a_n(g) q^n$$

is in $\mathcal{N}_{Q_g} \subset \mathcal{N}_Q$.

Lemma 4.9 *The map $g \mapsto g \otimes \xi_g$ defines a bijection between $\mathcal{N}'$ and $\mathcal{N}_Q$.*

Proof: Suppose we are given a newform f in $\mathcal{N}_Q$; i.e., f is in $\mathcal{N}_{\Gamma_0(Q)}$ and $\bar{\rho}_f \sim \bar{\rho}$. For $q \in Q$, we have (by lemma 2.44 for example) that

$$\rho_f|_{G_q} \sim \begin{pmatrix} \xi_{f,q} & 0 \\ 0 & \epsilon\xi_{f,q}^{-1} \end{pmatrix},$$

where $\xi_{f,q} : G_q \to K^\times$ is a character whose reduction is the unramified character sending Frob_q to α_q. Note that the characters $\xi_{f,q}|_{I_q}$ have ℓ-power order. In particular $N_f/N_\emptyset$ is Q_f^2 where Q_f is the product of the primes $q \in Q$ such that $\xi_{f,q}$ is ramified (theorem 3.1 (d)). Note also that there is a unique character

$$\xi_f : (\mathbb{Z}/Q_f\mathbb{Z})^\times \to K^\times$$

such that

$$I_q \to G_{\mathbb{Q}} \twoheadrightarrow \mathrm{Gal}\,(\mathbb{Q}(\zeta_{Q_f})/\mathbb{Q}) \xrightarrow{\xi_f} K^\times$$

coincides with the restriction of $\xi_{f,q}$ to I_q for each $q|Q_f$. (We have written ξ_f for the character of $G_{\mathbb{Q}}$ as well as the corresponding Dirichlet character. We shall also write ξ_f for the corresponding character of Δ_Q.) Let g denote the newform associated to the eigenform

$$\sum \xi_f^{-1}(n) a_n(f) q^n \in S_2(\Gamma_0(N_\emptyset) \cap \Gamma_1(Q_f^2), K).$$

By proposition 2.6 and theorem 3.1 we have

- $\rho_g \sim \rho_f \otimes \xi_f^{-1}$,
- $\psi_g = \xi_f^{-2}$,
- $N_g = N_\emptyset Q_f$ and
- $a_q(g) = (\xi_f|_{G_q}^{-1}\xi_{f,q})(\mathrm{Frob}_q)$ has reduction $\bar{\xi}_{f,q}(\mathrm{Frob}_q) = \alpha_q$ for all q dividing Q_f.

Therefore g is in $\mathcal{N}'$ and $f = g \otimes \xi_g$. In particular $g \mapsto g \otimes \xi_g$ is surjective and $\xi_g = \xi_{g \otimes \xi_g}$. Injectivity follows as well on noting that if $g \otimes \xi_g = g' \otimes \xi_{g'}$ then $\xi_g = \xi_{g'}$, hence $g = g'$. □

Recall that $\mathbb{T}_\mathcal{O}$ contains the operators $\langle d \rangle$ for d in $(\mathbb{Z}/M\mathbb{Z})^\times$, and note that $\langle d \rangle - 1 \in \mathfrak{m}$ for all d. Let H' denote the kernel of

$$(\mathbb{Z}/M\mathbb{Z})^\times \twoheadrightarrow (\mathbb{Z}/(\prod_{q \in Q} q)\mathbb{Z})^\times \twoheadrightarrow \Delta_Q.$$

Since the order of H' is not divisible by ℓ, we find that $\sum_{d \in H'} \langle d \rangle \notin \mathfrak{m}$. It follows that if $d \in H'$ then $\langle d \rangle = 1$ in $\mathbb{T}_\mathfrak{m}$. We may therefore regard $\mathbb{T}_\mathfrak{m}$ as an $\mathcal{O}[\Delta_Q]$-algebra via the map $d \mapsto \langle d \rangle$. Recall that $\mathbb{T}_Q$ is considered an $\mathcal{O}[\Delta_Q]$-algebra via the map

$$\mathcal{O}[\Delta_Q] \to R_Q \to \mathbb{T}_Q.$$

If p is a prime not in Q, let x_p denote the unique element of Δ_Q such that $x_p^{-2} = \bar{p}$.

Proposition 4.10 *There is an isomorphism $\mathbb{T}_Q \xrightarrow{\sim} \mathbb{T}_\mathfrak{m}$ of $\mathcal{O}[\Delta_Q]$-algebras such that $T_p \mapsto x_p T_p$ for all primes $p \notin \Sigma$ with $p \nmid \ell N(\bar{\rho})$.*

Proof: We may enlarge K so that we are in the setting of lemma 4.9. We then define an isomorphism of K-algebras

$$\kappa : \mathbb{T}_\mathfrak{m} \otimes K \xrightarrow{\sim} \prod_{g \in \mathcal{N}'} K$$

such that for each $g \in \mathcal{N}'$:

- $\kappa(T_p)_g = a_p(g)$ if p is not in Σ or if p divides Q_g,
- $\kappa(T_p)_g = 0$ if p is in $\Sigma - Q$,
- $\kappa(T_p)_g$ is the root of $X^2 - a_p(g) + \psi_g(p)p$ with reduction α_p if p is in Q but does not divide Q_g, and
- $\kappa(d)_g = \psi_g(d)$ for all $d \in \Delta_Q$.

The existence of such an isomorphism follows from lemma 4.4 which gives

$$\mathbb{T}_{\mathfrak{m}} \otimes K \xrightarrow{\sim} \prod_{g \in \mathcal{N}_\Gamma} \prod_{\mathfrak{p} \in \mathcal{M}_g} (A_{K,g}/I_{K,g})_{\mathfrak{p}},$$

where $\mathcal{M}_g$ is the set of prime ideals in $A_{K,g}/I_{K,g}$ whose preimage in $\mathbb{T}_{\mathcal{O}}$ is contained in $\mathfrak{m}$. If g is not in $\mathcal{N}'$, then $\mathcal{M}_g$ is empty. If g is in $\mathcal{N}'$, then $\mathcal{M}_g$ consists of the prime ideal corresponding to the eigenform whose eigenvalues are prescribed as above, and one checks that $(A_{K,g}/I_{K,g})_{\mathfrak{p}} = K$.

Viewing $\mathbb{T}_Q$ as a subalgebra of $\prod_{f \in \mathcal{N}_Q} K$ and matching indices via the bijection $f = g \otimes \xi_g \leftrightarrow g$, we obtain an injective homomorphism of $\mathcal{O}[\Delta_Q]$-algebras

$$\kappa' : \mathbb{T}_Q \longrightarrow \prod_{g \in \mathcal{N}'} K$$

such that

$$\kappa'(T_p) = (\xi_g(p) a_p(g))_g = \kappa(x_p T_p)$$

for primes $p \notin \Sigma$ such that $p \nmid \ell N_\emptyset$. Since $\mathbb{T}_Q$ is generated over $\mathcal{O}$ by the set of such T_p, we see that $\kappa'(\mathbb{T}_Q)$ is contained in $\kappa(\mathbb{T}_{\mathfrak{m}})$. On the other hand, the image of κ' contains the image of Δ_Q, hence contains $\kappa(\langle d \rangle)$ for d in $(\mathbb{Z}/M\mathbb{Z})^\times$ as well as $\kappa(T_p)$ for $p \notin \Sigma$ with $p \nmid N_\emptyset \ell$.

It remains to prove that $\kappa(T_p)$ is in the image of $\kappa'(\mathbb{T}_Q)$ for p dividing $N_\emptyset \ell \prod_{q \in Q} q$. Consider the composite

$$R_Q \xrightarrow{\phi_Q} \mathbb{T}_Q \xrightarrow{\kappa'} \prod_{g \in \mathcal{N}'} K.$$

The pushforward of ρ_Q^{univ} is equivalent to $\prod \rho_g \otimes \xi_g$ and therefore $\prod \rho_g$ is equivalent to the pushforward of

$$\rho_Q^{\mathrm{univ}} \otimes \xi_Q^{-1}$$

(where ξ_Q was defined in section 2.8). For primes p dividing $N_\emptyset$, we recover $\kappa(T_p)$ as the image of Frob_p on the I_p-coinvariants, hence $\kappa(T_p)$ is in $\kappa'(\mathbb{T}_Q)$. For $q \in Q$ and $g \in \mathcal{N}'$, the pushforward of $\xi_Q|_{G_q}^{-1} \cdot \xi_{q,Q}$ to the g-component is an unramified summand of $\rho_g|_{G_q}$ whose reduction sends Frob_q to α_q. It follows that this character maps Frob_q to $\kappa(T_q)$ and we conclude that $\kappa(T_q) \in \kappa'(\mathbb{T}_Q)$. Finally in the case that ℓ does not divide $N_\emptyset$, we appeal to lemma 4.1 to conclude that $\kappa(T_\ell) \in \kappa'(\mathbb{T}_Q)$. □

Auxiliary primes: For Σ as in proposition 4.10, i.e. a set of primes satisfying (4.2.2), a somewhat simpler argument provides a similar a description of $\mathbb{T}_\emptyset$ with Γ replaced by

$$\Gamma' = \Gamma_0(N_\emptyset) \cap \Gamma_H(M) \tag{4.2.5}$$

where H is the ℓ-Sylow subgroup of $(\mathbb{Z}/M\mathbb{Z})^\times$. While we shall make no direct use of this, the group Γ' will play a role in the proof of theorem 3.31 and we shall need to choose Σ so that Γ' has no elliptic elements; i.e., non-trivial elements of finite order. This is the case for example if Σ contains a prime $p \not\equiv 1 \bmod \ell$ with $p > 3$. The group Γ' is then contained in $\Gamma_1(N)$ for some integer $N > 3$, and therefore has no elliptic elements. Indeed if $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ has finite order then the roots of $X^2 - (a+d)X + 1$ are roots of unity and we deduce that this matrix is the identity.

In order to show that Σ can be so chosen, we appeal to the following lemma (cf. [DT2], lemma 3).

Lemma 4.11 *Suppose that G is a finite group, $\chi : G \to \bar{k}^\times$ is a character of order d and $\bar{\rho} : G \to GL_2(\bar{k})$ is a representation. Suppose that for all $g \notin \ker \chi$ we have*

$$(\operatorname{tr} \bar{\rho}(g))^2 / \det \bar{\rho}(g) = (1 + \chi(g))^2 / \chi(g). \tag{4.2.6}$$

- *If $d > 3$, then $\bar{\rho}$ is reducible.*
- *If $d = 3$, then $\bar{\rho}$ is reducible or has projective image isomorphic to A_4.*
- *If $d = 2$, then $\bar{\rho}|_{\ker \chi}$ is reducible.*

Proof: Note that if $\bar{\rho}(g)$ is a scalar then $\chi(G) = 1$. Hence χ induces a surjective homomorphism $\chi' : G' \to C_d$ where G' is the projective image of $\bar{\rho}$ and C_d is cyclic of order d. Furthermore if $d = 2$, then every element of $G' - \ker \chi'$ has order 2. The lemma then follows from theorem 2.47(b). □

4.3 Proof of theorem 3.31

We shall give a proof of theorem 3.31 which is based on the q-expansion principle rather than the method of de Shalit [dS] employed in [TW]. It will be more convenient to consider the action of the Hecke operators on the full space of modular forms $M_2(\Gamma)$. The Riemann-Roch theorem shows that the dimension of $M_2(\Gamma)$ is $g + s - 1$ where g is the genus of the modular curve X_Γ associated to Γ and s is the number of cusps on X_Γ (see for example [Shi2] thm. 2.23 or [DI] (12.1.5)).

Eisenstein maximal ideals: One can give an explicit description of a space of Eisenstein series $G_2(\Gamma)$ so that

$$M_2(\Gamma) = S_2(\Gamma) \oplus G_2(\Gamma).$$

(See for example [Hi3], lemma 5.2.)

There is a natural action on $M_2(\Gamma)$ by the Hecke operators $\langle d \rangle$ for all $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ and T_p for all primes p. We shall only need to use the operators $\langle d \rangle$ and T_p for p not dividing N, and we let $\tilde{\mathbb{T}}_{\mathbf{Z}}$ denote the subring of End $(M_2(\Gamma))$

these generate. The ring $\tilde{\mathbb{T}}_{\mathbf{Z}}$ is commutative and is finitely generated and free as a $\mathbb{Z}$-module. That $\tilde{\mathbb{T}}_{\mathbf{Z}}$ is finitely generated is proved for example by showing that $M_2(\Gamma, \mathbb{Z})$, the set of forms with integer Fourier coefficients at ∞, is stable under the Hecke operators and contains a basis for $M_2(\Gamma)$ (see [DI], cor. 12.3.12 and prop. 12.4.1). Alternatively, one can show that $\tilde{\mathbb{T}}_{\mathbf{Z}}$ acts faithfully on the cohomology of the non-compact modular curve Y_Γ (cf. section 1.3).

For any ring A, we write $M_2(\Gamma, A)$ for $M_2(\Gamma, \mathbb{Z}) \otimes A$ and regard this as a module for $\tilde{\mathbb{T}}_A := \tilde{\mathbb{T}}_{\mathbf{Z}} \otimes A$. If $\mathfrak{n}$ is a maximal ideal of $\tilde{\mathbb{T}}_{\mathcal{O}}$ we write $\tilde{\mathbb{T}}_{\mathfrak{n}}$ for the localization. If $\mathfrak{m}$ is a maximal ideal of $\mathbb{T}_{\mathcal{O}}$ then we let $\tilde{\mathfrak{m}}$ denote its preimage in $\tilde{\mathbb{T}}_{\mathcal{O}}$.

We say that a maximal ideal $\mathfrak{n}$ of $\tilde{\mathbb{T}}_{\mathcal{O}}$ is *Eisenstein* if

$$T_p \equiv p + 1 \bmod \mathfrak{n} \quad \text{for all } p \equiv 1 \bmod N.$$

One sees from the explicit description of the Eisenstein series that $T_p = p+1$ on $G_2(\Gamma)$ if $p \equiv 1 \bmod N$. We shall be interested in the *non-Eisenstein* maximal ideals because of the following lemma (see [R5], thm. 5.2(c)).

Lemma 4.12 *Suppose that ℓ is odd. The representation*

$$\rho_{\mathfrak{m}} : G_{\mathbf{Q}} \to GL_2(\mathbb{T}_{\mathcal{O}}/\mathfrak{m})$$

is absolutely irreducible if and only if $\tilde{\mathfrak{m}}$ is not Eisenstein.

Proof: If $\rho_{\mathfrak{m}}$ is not absolutely irreducible then $\rho_{\mathfrak{m}} \sim \chi_1 \oplus \chi_2$ with N divisible by the product of the conductors of χ_1 and χ_2. Furthermore if ℓ does not divide N, then one of the characters is unramified at ℓ while the other coincides with the cyclotomic character ϵ on I_ℓ. Therefore $\tilde{\mathfrak{m}}$ is Eisenstein. Conversely if $\tilde{\mathfrak{m}}$ is Eisenstein then proposition 2.6 implies that $\rho_{\mathfrak{m}}$ restricted to $G_{\mathbf{Q}(\zeta_{N\ell})}$ has trivial semisimplification from which it follows that $\rho_{\mathfrak{m}}$ is also reducible. □

Differentials: Suppose now that Σ is a finite set of primes satisfying (4.2.2) Moreover we assume that Σ contains a prime $p > 3$ with $p \not\equiv 1 \bmod \ell$. Let Γ and Γ' be defined as in (4.2.4) and (4.2.5); i.e.

$$\begin{aligned} \Gamma &= \Gamma_0(N_\emptyset) \cap \Gamma_1(M) \\ \Gamma' &= \Gamma_0(N_\emptyset) \cap \Gamma_H(M) \end{aligned}$$

where H is the ℓ-Sylow subgroup of $(\mathbb{Z}/M\mathbb{Z})^\times$. Let X and X' denote the modular curves associated to Γ and Γ'.

We first consider the case where $\ell \nmid N_\emptyset$. The curve X has a smooth proper model $\mathcal{X}$ over $\mathbb{Z}[1/N_\emptyset M]$ such that the complement of the cusps parametrizes cyclic isogenies $(E_1, i_1) \to (E_2, i_2)$ of degree $N_\emptyset$ where E_j is an elliptic curve and i_j is an embedding $\mu_M \hookrightarrow E_j$. (See [DR] or [Kat] for example.) The action of H on X extends to $\mathcal{X}$ and the quotient $\mathcal{X}' = \mathcal{X}/H$ is a smooth

model over $\mathbb{Z}[1/N_\emptyset M]$ for X'. The natural projection $\mathcal{X} \to \mathcal{X}'$ is étale. (The fact that it is étale on the complement of the cusps follows from the natural moduli-theoretic description of $\mathcal{X}'$ using the fact that $\Gamma_H(M) \subset \Gamma_1(p)$ for some $p > 3$. One then need only verify that $X \to X'$ is unramified at the cusps.)

For a $\mathbb{Z}[1/N_\emptyset M]$-algebra A, we define

- $\Omega_A = H^0(\mathcal{X}_A, \Omega^1_{\mathcal{X}_A/A})$,
- $\Omega'_A = H^0(\mathcal{X}'_A, \Omega^1_{\mathcal{X}'_A/A})$,
- $\tilde{\Omega}_A = H^0(\mathcal{X}_A, \Omega^1_{\mathcal{X}_A/A}(\mathcal{D}))$,
- $\tilde{\Omega}'_A = H^0(\mathcal{X}'_A, \Omega^1_{\mathcal{X}'_A/A}(\mathcal{D}'))$,

where $\mathcal{D}$ (resp. $\mathcal{D}'$) denotes the reduced divisor defined by the cusps of $\mathcal{X}$ (resp. $\mathcal{X}'$). The q-expansion principle and standard base-change arguments allow us to identify these with $S_2(\Gamma, A)$, $S_2(\Gamma', A)$, $M_2(\Gamma, A)$ and $M_2(\Gamma', A)$ (see [Kat], sec. 1.6, 1.7). We may therefore regard the first two of these as modules for $\mathbb{T}_A$ and all of them as modules for $\tilde{\mathbb{T}}_A$.

Lemma 4.13 *Suppose that $A = k$ or K.*

(a) $\tilde{\Omega}_\mathcal{O} \otimes_\mathcal{O} A \cong \tilde{\Omega}_A$ *and* $\tilde{\Omega}'_\mathcal{O} \otimes_\mathcal{O} A \cong \tilde{\Omega}'_A$.

(b) *The natural map $\mathcal{X} \to \mathcal{X}'$ induces an isomorphism*

$$\tilde{\Omega}'_A \xrightarrow{\sim} \tilde{\Omega}_A^H.$$

(c) *If $\mathfrak{n}$ is a non-Eisenstein maximal ideal of $\tilde{\mathbb{T}}_\mathcal{O}$, then the localization at $\mathfrak{n}$ of $\Omega_\mathcal{O} \to \tilde{\Omega}_\mathcal{O}$ is an isomorphism.*

(d) *If $\mathfrak{m}$ is a maximal ideal of $\mathbb{T}_A$ then $\Omega_A[\mathfrak{m}]$ is one-dimensional over $\mathbb{T}_A/\mathfrak{m}$.*

Sketch of proof:

(a) In the case $A = K$ this follows from the fact that K is flat over $\mathcal{O}$, so suppose $A = k$. Identify $\tilde{\Omega}_\mathcal{O}/\lambda\tilde{\Omega}_\mathcal{O}$ with the direct image of $\tilde{\Omega}_k$ under $\mathcal{X}_k \to \mathcal{X}_\mathcal{O}$ and use that $H^1(\mathcal{X}_k, \Omega^1_{\mathcal{X}_k/k}(\mathcal{D}))$ vanishes by Serre duality. The case of $\tilde{\Omega}'$ is similar. (See [Maz1] sec. II.3.)

(b) Using the fact that $\mathcal{X} \to \mathcal{X}'$ is étale one identifies the pull-back of $\Omega^1_{\mathcal{X}'_A/A}$ with $\Omega^1_{\mathcal{X}_A/A}$ and that of $\mathcal{D}'$ with $\mathcal{D}$.

(c) The map is injective and one proves that its cokernel is free and annihilated by the operators $T_p - (p+1)$ for all $p \equiv 1 \bmod N_\emptyset M$. To prove the latter assertion, first observe that it holds with $\mathcal{O}$ replaced by $\mathbb{C}$, then by $\mathbb{Z}[1/N_\emptyset M]$.

(d) This follows from

$$\Omega_A[\mathfrak{m}] \cong S_2(\Gamma, A)[\mathfrak{m}] \cong \mathrm{Hom}_A(\mathbb{T}_A, A)[\mathfrak{m}] \cong \mathrm{Hom}_A(\mathbb{T}_A/\mathfrak{m}, A),$$

where the middle isomorphism is that of proposition 1.34, but let us reformulate the argument in a way more easily generalized to the case of ℓ dividing $N_\emptyset$ discussed below. Note that $\Omega_A[\mathfrak{m}]$ is non-zero since $\mathbb{T}_\mathfrak{m}$ acts faithfully on $\Omega_\mathfrak{m}$. To prove that the dimension is at most one we can enlarge the field A and assume $\mathbb{T}_A/\mathfrak{m} \cong A$. One then shows that an eigenform f in Ω_A (for all the Hecke operators T_p and $\langle d \rangle$) is determined by its eigenvalues and the first coefficient of its q-expansion. This follows from the fact that f is determined by its q-expansion and for all n, $a_n(f) = a_1(T_n f)$ and T_n can be expressed in terms of the T_p and $\langle d \rangle$. (See [Maz1], sec. II.9.) □

We now explain how the situation changes if ℓ divides $N_\emptyset$. In that case X has a regular model $\mathcal{X}$ over $\mathbb{Z}[\ell/N_\emptyset M]$ with the same moduli-theoretic description as above. This model is smooth over $\mathbb{Z}[1/N_\emptyset M]$, but $\mathcal{X}_{\mathbf{F}_\ell}$ has two smooth irreducible components crossing at ordinary double points as in [DR] sec. V.1. The quotient $\mathcal{X}' = \mathcal{X}/H$ is a regular model for X' over $\mathbb{Z}[\ell/N_\emptyset M]$ and $\mathcal{X} \to \mathcal{X}'$ is étale. For a $\mathbb{Z}[\ell/N_\emptyset M]$-algebra A we define Ω_A, Ω'_A, $\tilde{\Omega}_A$ and $\tilde{\Omega}'_A$ as before, except that Ω^1 is replaced by the sheaf of regular differentials (see [DR] sec. I.2, [Maz1] sec. II.6 or [MRi] sec. 7). Formation of these modules again commutes with change of the base A, but we can no longer identify them with $S_2(\Gamma, A)$, $S_2(\Gamma', A)$, $M_2(\Gamma, A)$ and $M_2(\Gamma', A)$ if ℓ is not invertible in A. For every A, there is a natural action of $\mathbb{T}_A$ (resp. $\tilde{\mathbb{T}}_A$) on Ω_A and Ω'_A (resp. $\tilde{\Omega}_A$ and $\tilde{\Omega}'_A$). In the case of $\mathbb{T}_A$ this is proved by identifying Ω_A (resp. Ω'_A) with the cotangent space at the origin for the Néron model over A for the Jacobian of X (resp. X'). In the case of $\tilde{\mathbb{T}}_A$ the action is defined using Grothendieck-Serre duality (as in [Maz1] sec. II.6 or [MRi] sec. 7).

Lemma 4.14 *If ℓ divides $N_\emptyset$ then lemma 4.13 carries over with the above notation and the additional hypothesis $T_\ell \notin \mathfrak{m}$ for part* (d).

The proof is essentially the same except that part (d) is more delicate in the case $A = k$. For the proof in that case we refer the reader to [W3], lemma 2.2 (from which the result stated here is immediate). We remark however that we shall only use lemma 4.14 if $\rho_\mathfrak{m}$ is not good. In that case one can also deduce the statement here from the argument used in the proof of [MRi] prop. 20, which shows instead that $\dim_k \Omega_k[\mathfrak{m}'] = 1$ where $\mathfrak{m}'$ is a certain maximal ideal corresponding to $\mathfrak{m}$, but contained in an algebra defined using the operators $\langle d \rangle$, T_p for $p \neq \ell$ and w_ℓ.

Remark 4.15 The approach in [TW] to proving theorem 3.31 is based on the result that under certain hypotheses $(\mathcal{T}_\ell(J) \otimes_{\mathbb{Z}_\ell} \mathcal{O})_\mathfrak{m}$ is free of rank two

over $\mathbb{T}_{\mathfrak{m}}$ (see [W3], thm. 2.1 and its corollaries). This result generalizes work of Mazur Ribet, and Edixhoven (see secs. 14 and 15 of ch. II of [Maz1], thm. 5.2 of [R5], [MRi] and sec. 9 of [Edi]), and the key to its proof is lemmas 4.13(d) and 4.14. We shall instead give a proof of theorem 3.31 which is based directly on lemma 4.13, (d).

The fact that $(\mathcal{T}_\ell(J)\otimes_{\mathbf{Z}_\ell}\mathcal{O})_{\mathfrak{m}}$ is free of rank two over $\mathbb{T}_{\mathfrak{m}}$ actually underlies Wiles' approach to many intermediate results along the way to proving the Shimura-Taniyama conjecture for semistable elliptic curves. We shall appeal to a special case in the course of proving theorem 3.36 below.

Proof of the theorem: Suppose we are given a representation $\bar{\rho}$ as in section 3.3 and a set of primes Q as in the statement of theorem 3.31. We apply lemma 4.11 with $G = G_{\mathbb{Q}}$ and $\chi = \epsilon$ to choose an auxiliary prime p not dividing $6N_\emptyset \ell \prod_{q\in Q} q$ such that

- $p \not\equiv 1 \bmod \ell$
- $\mathrm{tr}\,(\bar{\rho}(\mathrm{Frob}_p))^2 \neq (p+1)^2$.

To prove theorem 3.31 we may replace K by a larger field and assume that k contains the eigenvalues of $\bar{\rho}(\mathrm{Frob}_p)$.

Let $\Sigma = Q \cup \{p\}$. Thus Σ is a set of primes as in proposition 4.10 and Σ contains a prime $p > 3$ such that $p \not\equiv 1 \bmod \ell$. We let Γ be as in (4.2.4); thus

$$\Gamma = \Gamma_0(N_\emptyset) \cap \Gamma_1(M)$$

where $M = p^2 \prod_{q\in Q} q$, and we choose the maximal ideal $\mathfrak{m}$ of $\mathbb{T}_{\mathcal{O}}$ as in proposition 4.10. Let H denote the ℓ-Sylow subgroup of $(\mathbb{Z}/M\mathbb{Z})^\times$ and regard $\mathbb{T}_{\mathcal{O}}$ and hence $\mathbb{T}_{\mathfrak{m}}$ as an $\mathcal{O}[H]$-algebra via $d \mapsto \langle d\rangle$. In view of proposition 4.10, theorem 3.31 is equivalent to the following:

Theorem 4.16 *$\mathbb{T}_{\mathfrak{m}}$ is free over $\mathcal{O}[H]$.*

Proof: Consider the $\tilde{\mathbb{T}}_{\mathcal{O}}$-module

$$\tilde{L} = \mathrm{Hom}_{\mathcal{O}}(\tilde{\Omega}_{\mathcal{O}}, \mathcal{O})$$

and the $\mathbb{T}_{\mathcal{O}}$-module

$$L = \mathrm{Hom}_{\mathcal{O}}(\Omega_{\mathcal{O}}, \mathcal{O}).$$

Since $X \to X'$ is unramified, the Riemann-Hurwitz formula implies that

$$\dim M_2(\Gamma) = g + s - 1 = \#H(g' + s' - 1) = \#H \dim M_2(\Gamma'),$$

where g (resp. g') is the genus and s (resp. s') is the number of cusps of X (resp. X'). It follows from lemma 4.13 (a) (and lemma 4.14) that

$$\dim_K(\tilde{L} \otimes_{\mathcal{O}} K) = \#H \dim_k \tilde{\Omega}'_k \qquad (4.3.1)$$

On the other hand, by (b) we have that

$$\tilde{L}/(\mathfrak{a},\lambda)\tilde{L} \cong \mathrm{Hom}_k(\tilde{\Omega}_k^H, k) \cong \mathrm{Hom}_k(\tilde{\Omega}'_k, k) \tag{4.3.2}$$

where $\mathfrak{a} = \ker(\mathcal{O}[H] \to \mathcal{O})$ is the augmentation ideal. By Nakayama's lemma and (4.3.2), $\tilde{L}$ is generated as an $\mathcal{O}[H]$-module by d elements where $d = \dim_k \tilde{\Omega}'_k$. By (4.3.1) any surjective homomorphism

$$\mathcal{O}[H]^d \to \tilde{L}$$

is in fact an isomorphism. Hence $\tilde{L}$ is free over $\mathcal{O}[H]$, as is

$$L_{\mathfrak{n}} = \mathrm{Hom}_{\mathcal{O}}(\Omega_{\mathfrak{n}}, \mathcal{O}) \cong \mathrm{Hom}_{\mathcal{O}}(\tilde{\Omega}_{\mathfrak{n}}, \mathcal{O})$$

for each non-Eisenstein maximal ideal $\mathfrak{n}$ of $\tilde{\mathbb{T}}_{\mathcal{O}}$ (by part (c) of lemmas 4.13 and 4.14).

Since $\rho_{\mathfrak{m}} \sim \bar{\rho}$ is absolutely irreducible, $\tilde{\mathfrak{m}}$ is not Eisenstein (by lemma 4.12), and it follows that $L_{\mathfrak{m}}$ is free over $\mathcal{O}[H]$.

Since $L \otimes_{\mathcal{O}} K$ is free of rank one over $\mathbb{T}_K$ (lemma 1.34), we have that $L_{\mathfrak{m}} \otimes_{\mathcal{O}} K$ is free of rank one over $\mathbb{T}_{\mathfrak{m}} \otimes_{\mathcal{O}} K$. If $\ell \nmid N_{\emptyset}$ then lemma 4.13 implies that

$$L/\mathfrak{m}L \cong \mathrm{Hom}_k(\Omega_k[\mathfrak{m}], k)$$

is one-dimensional over k. The same assertion holds if $\ell | N_{\emptyset}$ by lemma 4.14 (from the definition of $\mathfrak{m}$ in this case, we see that it does not contain T_ℓ). It follows from Nakayama's lemma that $L_{\mathfrak{m}}$ is free of rank one over $\mathbb{T}_{\mathfrak{m}}$. Therefore $\mathbb{T}_{\mathfrak{m}}$ is free over $\mathcal{O}[H]$. □

4.4 Proof of theorem 3.36

Our proof of theorem 3.36 is based on Wiles' arguments in ch. 2 of [W3], which in turn are based on a method of Ribet [R4]. We shall reformulate the proof somewhat to underscore the relationship observed by Doi and Hida [Hi1] between the size of $\mathcal{O}/\eta$ and the value of an L-function, but after doing so we shall also sketch the more direct argument used by Wiles. Two important ingredients appear in both versions of the argument, and we shall discuss their proofs in section 4.5. These ingredients are

- the generalization of a result of Ihara [Ih] used in Ribet's argument;
- the generalization of a result of Mazur [Maz1] on the structure of the Tate module $T_\ell(J_0(N))$ as a module for $\mathbb{T}_{\mathbf{Z}_\ell}$.

Assume now that we are in the setting of theorem 3.36. In particular, $\bar{\rho}$ is a representation as in section 3.3, Σ is a finite set of primes contained in $\Sigma_{\bar{\rho}}$ and π is an $\mathcal{O}$-algebra homomorphism $\mathbb{T}_\Sigma \to \mathcal{O}$ arising from a newform f in $\mathcal{N}_\Sigma$ with coefficients in $\mathcal{O}$.

The symmetric square L-function: For each prime p let $\alpha_p(f)$ and $\beta_p(f)$ denote the roots of the polynomial $X^2 - a_p(f)X + \delta_p p = 0$, where $\delta_p = 0$ or 1 according to whether p divides N_f. Thus we have

$$L_p(f,s) = (1 - \alpha_p(f)p^{-s})^{-1}(1 - \beta_p(f)p^{-s})^{-1}.$$

If $p = \ell$ divides N/N_f, then we require that $\alpha_\ell(f)$ be the root which is a unit in $\mathcal{O}$, i.e. $a_\ell(g)$. We define the symmetric square L-function associated to f as

$$L(\mathrm{Symm}^2 f, s) = \prod_p L_p(\mathrm{Symm}^2 f, s),$$

where

$$L_p(\mathrm{Symm}^2 f, s) = (1 - \alpha_p^2(f)p^{-s})^{-1}(1 - \alpha_p(f)\beta_p(f)p^{-s})^{-1}(1 - \beta_p^2(f)p^{-s})^{-1}.$$

(We caution the reader we have defined the Euler factors rather naively at primes p such that p^2 divides N_f.) The product converges absolutely for real part of $s > 2$ and can be analytically continued to an entire function.

The calculation of η_Σ: We write $\wp$ for the kernel of π and I for the annihilator of $\wp$ in $\mathbb{T}_\Sigma$. Since $\mathbb{T}_\Sigma$ is reduced, we have $\wp \cap I = 0$ and $\wp \oplus I$ has finite index in $\mathbb{T}_\Sigma$. Note that

$$\mathcal{O}/\eta_\Sigma \cong \mathbb{T}_\Sigma/(\wp \oplus I).$$

Suppose that we are given a $\mathbb{T}_\Sigma$-module L which is finitely generated and free over $\mathcal{O}$, and such that $L \otimes K$ is free of rank d over $\mathbb{T}_\Sigma \otimes_\mathcal{O} K$. Suppose also that L is endowed with a perfect $\mathcal{O}$-bilinear pairing

$$\begin{array}{ccc} L \times L & \to & \mathcal{O} \\ (x,y) & \mapsto & \langle x,y \rangle \end{array}$$

such that $\langle Tx, y\rangle = \langle x, Ty\rangle$ for all $x, y \in L$ and $T \in \mathbb{T}_\Sigma$. Note that to give such a pairing is equivalent to giving an isomorphism

$$\begin{array}{ccc} L & \xrightarrow{\sim} & \mathrm{Hom}_\mathcal{O}(L, \mathcal{O}) \\ x & \mapsto & \langle x, \cdot \rangle \end{array}$$

of $\mathbb{T}_\Sigma$-modules. We shall refer to the module L together with the pairing $\langle\ ,\ \rangle$ as a *self-dual* $\mathbb{T}_\Sigma$-module of *rank d*. We then can give a lower bound for the size of $\mathcal{O}/\eta_\Sigma$ in terms of a basis $\{x_1, x_2, \ldots, x_d\}$ for the free $\mathcal{O}$-module $L[\wp]$:

Lemma 4.17 *We have*

$$\eta_\Sigma^d \subset \mathcal{O} \det(\langle x_i, x_j\rangle)_{i,j},$$

and equality holds if L is free over $\mathbb{T}_\Sigma$.

Proof: The modules $L[\wp]$ and $L/L[I]$ are free of rank d over $\mathcal{O}$ and the pairing $\langle\ ,\ \rangle$ induces an isomorphism

$$L/L[I] \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}}(L[\wp], \mathcal{O}). \tag{4.4.1}$$

The $\mathcal{O}$-module $M = L/(L[\wp] \oplus L[I])$ is annihilated by η_Σ and is generated by d elements. Furthermore M is isomorphic to $(\mathcal{O}/\eta_\Sigma)^d$ if L is free over $\mathbb{T}_\Sigma$. It follows that

$$\#(\mathcal{O}/\eta_\Sigma)^d \geq \#M,$$

with equality if L is free. The cardinality of M is that of the cokernel of the map

$$L[\wp] \hookrightarrow \mathrm{Hom}_{\mathcal{O}}(L[\wp], \mathcal{O})$$

arising from the pairing $\langle\ ,\ \rangle$, and this is precisely

$$\mathcal{O}/\det(\langle x_i, x_j\rangle)_{i,j}.$$

□

Recall that proposition 4.7 establishes an isomorphism between $\mathbb{T}_\Sigma$ and $\mathbb{T}_{\mathfrak{m}}$, the localization at a certain maximal ideal $\mathfrak{m}$ of $\mathbb{T}_{\mathcal{O}} = \mathbb{T}_{\mathbf{Z}} \otimes \mathcal{O}$, where

$$\mathbb{T}_{\mathbf{Z}} \subset \mathrm{End}\,(S_2(\dot{\Gamma}_0(N))),$$

and where $N = N_\Sigma$ is defined in (4.2.1). Write X for $X_0(N)$ and J for $J_0(N)$. Recall that $H_1(X, \mathbb{Z})$ is endowed with the structure of a $\mathbb{T}_{\mathbf{Z}}$-module and

$$\mathcal{T}_\ell(J) \otimes_{\mathbf{Z}_\ell} \mathcal{O} \cong H_1(X, \mathbb{Z}) \otimes \mathcal{O}$$

with that of a $\mathbb{T}_{\mathcal{O}}$-module. We also regard

$$H^1(X, \mathcal{O}) \cong \mathrm{Hom}(H_1(X, \mathbb{Z}), \mathcal{O})$$

as a $\mathbb{T}_{\mathcal{O}}$-module; as such it is naturally isomorphic to

$$\mathrm{Hom}_{\mathcal{O}}(\mathcal{T}_\ell(J) \otimes_{\mathbf{Z}_\ell} \mathcal{O}, \mathcal{O}).$$

The Weil pairing defines a perfect pairing

$$\mathcal{T}_\ell(J) \times \mathcal{T}_\ell(J) \to \mathbb{Z}_\ell$$

(with $(e^{2\pi i/\ell^n})_n$ as the chosen generator for $\varprojlim \mu_{\ell^n}(\mathbb{C})$). The composite

$$\mathcal{T}_\ell(J) \xrightarrow{w_*} \mathcal{T}_\ell(J) \to \mathrm{Hom}_{\mathbf{Z}_\ell}(\mathcal{T}_\ell(J), \mathbb{Z}_\ell) \tag{4.4.2}$$

is an isomorphism of $\mathbb{T}_{\mathbf{Z}_\ell}$-modules where w_* is induced by the involution $w = w_N$ of X (see section 1.4). Tensoring with $\mathcal{O}$ and localizing at $\mathfrak{m}$, we regard

$$L_{\mathcal{T}} = (\mathcal{T}_\ell(J) \otimes_{\mathbf{Z}_\ell} \mathcal{O})_{\mathfrak{m}}$$

as a self-dual $\mathbb{T}_\Sigma$-module of rank 2 via the isomorphism $\mathbb{T}_\Sigma \cong \mathbb{T}_\mathfrak{m}$ of proposition 4.7. We shall write $\langle , \rangle_\mathcal{T}$ for the pairing obtained from (4.4.2). Similarly, using w^*, the cup product and Poincaré duality, we regard

$$L_H = H^1(X, \mathcal{O})_\mathfrak{m}$$

as a self-dual $\mathbb{T}_\Sigma$-module of rank 2, with $\langle x, y \rangle_H$ defined by the image of $x \cup w^* y$ under the canonical isomorphism of $H^2(X, \mathcal{O})$ with $\mathcal{O}$. Note that L_H is canonically isomorphic as a $\mathbb{T}_\Sigma$-module to $\mathrm{Hom}_\mathcal{O}(L_\mathcal{T}, \mathcal{O})$. In the next section we shall discuss the following generalization of a result of Mazur [Maz1].

Theorem 4.18 *The $\mathbb{T}_\Sigma$-modules $L_\mathcal{T}$ and L_H are free.*

Corollary 4.19 *If $\{x, y\}$ is a basis for $L_\mathcal{T}[\wp]$ (resp. $L_H[\wp]$), then η_Σ is generated by $\langle x, y \rangle_\mathcal{T}$ (resp. $\langle x, y \rangle_H$).*

This follows from lemma 4.17, theorem 4.18 and skew-symmetry of the pairings.

Hida's formula: We now explain how the value of $\langle x, y \rangle_H$ in corollary 4.19 is related to $L(\mathrm{Symm}^2 f, 2)$ by a formula of Hida (see [Hi1] sec. 5, [W3] sec. 4.1).

Recall that the homomorphism

$$\mathbb{T}_\mathcal{O} \to \mathbb{T}_\mathfrak{m} \cong \mathbb{T}_\Sigma \xrightarrow{\pi} \mathcal{O}$$

arises as $T_n \mapsto a_n(g)$ for a normalized eigenform g in $S_2(\Gamma_0(N), K)$ whose associated newform is f. We shall write $\mathcal{P}$ for the kernel of this homomorphism; thus $\mathcal{P}$ is the preimage in $\mathbb{T}_\mathcal{O}$ of the ideal $\wp$ of $\mathbb{T}_\Sigma$. Note that $\mathcal{P}_\mathfrak{m}$ corresponds to $\wp$ under the isomorphism $\mathbb{T}_\mathfrak{m} \cong \mathbb{T}_\Sigma$.

Choose a number field K_0 containing the Fourier coefficients of g and let $\mathcal{O}_0$ denote the valuation ring $\mathcal{O} \cap K_0$. We choose the basis $\{x, y\}$ for $L_H[\wp]$ in the image of

$$H^1(X, \mathcal{O}_0)[\mathcal{P}_0] \subset H^1(X, \mathcal{O})[\mathcal{P}] \xrightarrow{\sim} H^1(X, \mathcal{O})_\mathfrak{m}[\mathcal{P}] = L_H[\wp],$$

where $\mathcal{P}_0 = \mathcal{P} \cap \mathbb{T}_{\mathcal{O}_0}$. Let $\mathcal{P}_\mathbb{C}$ denote the kernel of the homomorphism

$$\theta_g : \mathbb{T}_\mathbb{C} \to \mathbb{C}$$

associated to the eigenform g. The two-dimensional complex vector space

$$(H^1(X, \mathcal{O}_0)[\mathcal{P}_0]) \otimes_{\mathcal{O}_0} \mathbb{C} = H^1(X, \mathbb{C})[\mathcal{P}_\mathbb{C}]$$

has a canonical basis

$$\{\omega_g, \bar{\omega}_{g^c}\},$$

where we view these differential forms as cohomology classes and the superscript c indicates complex conjugation applied to the Fourier coefficients. (Recall that ω_g is the holomorphic differential on X defined by $\sum a_n(g)q^{n-1}dq$, so $\bar{\omega}_{g^c}$ is the antiholomorphic differential defined by $\sum a_n(g)\bar{q}^{n-1}d\bar{q}$ where $q = e^{2\pi i\tau}$.) Let A denote the matrix in $GL_2(\mathbb{C})$ such that

$$(\omega_g, \bar{\omega}_{g^c}) = (x, y)A.$$

Theorem 4.20 *With the above notation we have (up to sign)*

$$\langle x, y\rangle = (i\pi \det A)^{-1} N_f L(\mathrm{Symm}^2 f, 2) \prod_{p|N/N_f} c_p(f)),$$

where

$$c_p(f) = \begin{cases} p^{\delta_p} L_p(\mathrm{Symm}^2 f, 2)^{-1} & \text{if } p \neq \ell \\ (\alpha_\ell(f) - \beta_\ell(f))(1 - \alpha_\ell^{-2}(f)) & \text{if } p = \ell. \end{cases}$$

Proof: We have

$$\langle x, y\rangle_H \det A = \int_X \omega_g \wedge \bar{\omega}_{wg^c} = 8\pi^2 i \langle g, (wg)^c\rangle$$

where the last $\langle , \rangle$ denotes the Petersson inner product and we have used that $wg^c = (wg)^c$ for forms on $\Gamma_0(N)$. We then appeal to a formula of Shimura (see [Shi5] (2.5) and [Hi1] (5.13)) to obtain

$$\langle g, (wg)^c\rangle = (48\pi)^{-1}[SL_2(\mathbb{Z}) : \Gamma_0(N)]\mathrm{res}_{s=2} D(g, wg, s).$$

By [Shi5] lemma 1, the Dirichlet series $D(g, h, s)$ is defined by $\sum a_n(g)a_n(h)n^{-s}$ and if g and h are normalized eigenforms then this has an Euler product expression in which the factors are

$$(1 - \alpha_p(g)\beta_p(g)\alpha_p(h)\beta_p(h)p^{-2s})(1 - \alpha_p(g)\alpha_p(h)p^{-s})^{-1} \cdot$$
$$(1 - \alpha_p(g)\beta_p(h)p^{-s})^{-1}(1 - \beta_p(g)\alpha_p(h)p^{-s})^{-1}(1 - \beta_p(g)\beta_p(h)p^{-s})^{-1},$$

(where the α_p's and β_p's are defined as they were for f). Using the recipe for obtaining g from f (see lemma 4.6) we find that if ℓ is not in Σ then

$$D(g, wg, s) = \pm D_\Sigma(f, f, s) \prod_{p|N/N_f} p^{-1},$$

(where the subscript Σ indicates that the primes in Σ are removed from the defining Euler product). If ℓ is in Σ, then we must also multiply by the factor

- $(1 - \ell^{-s})^{-1}$ if ℓ divides N_f;
- $(1 - \ell^{1-s})^{-1}(1 - \alpha_\ell^2(f)\ell^{-s})^{-1}(\beta_\ell(f) - \alpha_\ell^{2-s}(f))$ otherwise.

To obtain the formula in this last case, consider the eigenform

$$f_\alpha = f - \beta_\ell(f) f(\ell\tau)$$

on $\Gamma_0(N_f\ell)$. Since

$$w_{N_f\ell} f_\alpha = \pm(\ell f(\ell\tau) - \alpha_\ell^{-1}(f) f)$$

and

$$D(f_\alpha, \ell f(\ell\tau), s) = \alpha_\ell(f)\ell^{1-s} D(f_\alpha, f, s),$$

we get

$$D(f_\alpha, w_{N_f\ell} f_\alpha, s) = \pm\ell^{-1}(\beta_\ell(f) - \alpha_\ell(f)\ell^{2-s}) D(f_\alpha, f, s)$$

and use the Euler product. The Euler product expression also gives

$$D_\Sigma(f, f, s) = \frac{\zeta_N(s-1) L_\Sigma(\mathrm{Symm}^2 f, s)}{\zeta_N(2s-2)},$$

where $\zeta_N(s)$ is the Riemann zeta function with the Euler factors removed at primes dividing N. Thus

$$\mathrm{res}_{s=2} D(g, wg, s) = \pm\frac{6}{\pi^2} \frac{N L_\Sigma(\mathrm{Symm}^2 f, 2) \prod_{p|N/N_f} p}{[SL_2(\mathbb{Z}) : \Gamma_0(N)]},$$

with the extra factor of $(1-\ell^{-2})^{-1}$ or

$$(\beta_\ell(f) - \alpha_\ell(f))(1 - \beta_\ell^{-2}(f))^{-1}(1-\ell^{-1})^{-1}$$

if ℓ is in Σ. The theorem follows. □

Corollary 4.21 *With the above notation (in particular, $\Sigma \subset \Sigma_{\bar{\rho}}$), we have*

$$\frac{\ell^m N_f L_\Sigma(\mathrm{Symm}^2 f, 2)}{i\pi \det A} \in \mathcal{O}_0$$

is a generator of η_Σ, where

$$m = \begin{cases} 0, \text{ if } \ell \notin \Sigma, \\ 2, \text{ if } \ell \in \Sigma \text{ and } \ell | N_f, \\ 3, \text{ if } \ell \in \Sigma \text{ and } \ell \nmid N_f. \end{cases}$$

Remark 4.22 Without theorem 4.18 we obtain instead that η_Σ is contained in the ideal generated by the expression in the corollary.

Comparing η_Σ's: Suppose that $\Sigma' \subset \Sigma_{\bar{\rho}}$ is a finite set of primes containing Σ. In view of corollary 4.21, theorem 3.36 reduces to a comparison of $\det A$

and $\det A'$ where A' is defined using Σ' instead of Σ. We shall make the desired comparison using Wiles' generalization of a result of Ihara.

Applying proposition 4.7 to Σ', we have an isomorphism between $\mathbb{T}_{\Sigma'}$ and $\mathbb{T}'_{\mathfrak{m}'}$, where $\mathfrak{m}'$ is a certain maximal ideal of $\mathbb{T}'_{\mathcal{O}} = \mathbb{T}'_{\mathbf{Z}} \otimes \mathcal{O}$ with $\mathbb{T}'_{\mathbf{Z}}$ acting on $S_2(\Gamma_0(N_{\Sigma'}))$. We write $\mathcal{P}'$ for the preimage in $\mathbb{T}'_{\mathcal{O}}$ of $\wp'$. We let $N' = N_{\Sigma'}$, $X' = X_0(N')$, $J' = J_0(N')$ and define self-dual $\mathbb{T}'_{\mathfrak{m}'}$-modules $L'_{\mathcal{T}}$ and L'_H as above.

We now define a homomorphism

$$\phi_{\mathcal{T}} : L'_{\mathcal{T}} \to L_{\mathcal{T}}$$

of $\mathbb{T}'_{\mathfrak{m}'}$-modules in the case that $\Sigma' = \Sigma \cup \{p\}$ for some prime $p \in \Sigma_{\bar{p}} - \Sigma$. We first suppose that $p \neq \ell$, in which case $N' = Np^2$ and the surjective homomorphism of $\mathcal{O}$-algebras $\mathbb{T}'_{\mathfrak{m}'} \to \mathbb{T}_{\mathfrak{m}}$ (corresponding to $\mathbb{T}_{\Sigma'} \mapsto \mathbb{T}_{\Sigma}$) is defined by $T_r \mapsto T_r$ if $r \neq p$ and $T_p \mapsto 0$. The maps $\tau \mapsto \tau$, $\tau \mapsto p\tau$ and $\tau \mapsto p^2\tau$ on the upper half-plane induce maps $X' \to X$ which we denote α, β and γ. These give rise by functoriality to a map

$$\begin{array}{ccc} J' & \to & J \times J \times J \\ x & \mapsto & (\alpha_* x, \beta_* x, \gamma_* x) \end{array}$$

which induces an $\mathcal{O}$-linear homomorphism

$$\mathcal{T}_\ell(J') \otimes_{\mathbf{Z}_\ell} \mathcal{O} \to (\mathcal{T}_\ell(J) \otimes_{\mathbf{Z}_\ell} \mathcal{O})^3.$$

We compose with the projection to

$$L^3_{\mathcal{T}} = (\mathcal{T}_\ell(J) \otimes_{\mathbf{Z}_\ell} \mathcal{O})^3_{\mathfrak{m}}$$

followed by the map $L^3_{\mathcal{T}} \to L_{\mathcal{T}}$ defined by the matrix $(1, -p^{-1}T_p, p^{-1})$. The reason for this choice of matrix is that

$$g' = (\alpha^* - p^{-1}\beta^* T_p + p^{-1}\gamma^*)g \tag{4.4.3}$$

is the eigenform corresponding to $\mathcal{P}'$.

Lemma 4.23 *The composite*

$$\mathcal{T}_\ell(J') \otimes_{\mathbf{Z}_\ell} \mathcal{O} \to L_{\mathcal{T}} \tag{4.4.4}$$

is $\mathbb{T}'_{\mathcal{O}}$-linear where $L_{\mathcal{T}}$ is regarded as a $\mathbb{T}'_{\mathcal{O}}$-module via the surjection

$$\mathbb{T}'_{\mathcal{O}} \to \mathbb{T}'_{\mathfrak{m}'} \to \mathbb{T}_{\mathfrak{m}}.$$

We leave the proof to the reader after pointing out that the map on Hecke algebras can be rewritten as

$$\begin{array}{ccccccc} \mathbb{T}'_{\mathcal{O}} & \to & \mathbb{T}_{\mathcal{O}}[u]/(f(u)) & \to & \mathbb{T}_{\mathfrak{m}}[u]/(f(u)) & \to & \mathbb{T}_{\mathfrak{m}} \\ T_p & \mapsto & u & & u & \mapsto & 0, \end{array}$$

where $f(u) = u^3 - T_p u^2 + pu$.

It follows that (4.4.4) factors through $L'_{\mathcal{T}}$ and we thus obtain the desired homomorphism

$$\phi_{\mathcal{T}} : L'_{\mathcal{T}} \to L_{\mathcal{T}}$$

of $\mathbb{T}'_{\mathfrak{m}'}$-modules. Note that $\phi_{\mathcal{T}}$ is also the composite of the canonical splitting

$$L'_{\mathcal{T}} \hookrightarrow T_\ell(J') \otimes_{\mathbb{Z}_\ell} \mathcal{O}$$

with (4.4.4).

The construction of $\phi_{\mathcal{T}}$ is similar for $p = \ell$ except that we have only two copies of J and the map $L^2_{\mathcal{T}} \to L_{\mathcal{T}}$ is given by the matrix $(1, -\alpha_\ell^{-1})$ where α_ℓ is the unit root of (3.3.2). For arbitrary Σ' satisfying $\Sigma \subset \Sigma' \subset \Sigma_{\bar{p}}$ we define $\phi_{\mathcal{T}}$ as a composite of the maps defined above. It is independent of the choice of ordering of $\Sigma' - \Sigma$. Recall that $\phi'_{\mathcal{T}}$ is used to denote the dual map $L_{\mathcal{T}} \to L'_{\mathcal{T}}$. We define $\phi_H : L'_H \to L_H$ as the adjoint of the map ϕ'_H which renders the diagram

$$\begin{array}{ccc} \phi'_H : & H^1(X, \mathcal{O})_{\mathfrak{m}} & \to H^1(X', \mathcal{O})_{\mathfrak{m}'} \\ & \downarrow & \downarrow \\ & \mathrm{Hom}_{\mathcal{O}}(L_{\mathcal{T}}, \mathcal{O}) & \to \mathrm{Hom}_{\mathcal{O}}(L'_{\mathcal{T}}, \mathcal{O}) \end{array}$$

commutative (where the vertical maps are the natural isomorphisms and the bottom one is dual to $\phi_{\mathcal{T}}$).

The second crucial result whose discussion we postpone until the next section is the following generalization of a lemma of Ihara.

Lemma 4.24 *$\phi_{\mathcal{T}}$ and ϕ_H are surjective.*

Suppose again we are in the case $\Sigma' = \Sigma \cup \{p\}$ with $p \neq \ell$. One need only unravel the definition of ϕ'_H to see that the diagram

$$\begin{array}{ccc} H^1(X, \mathcal{O}_0)[\mathcal{P}_0] & \to & H^1(X', \mathcal{O}_0)[\mathcal{P}'_0] \\ \downarrow & & \downarrow \\ L_H & \xrightarrow{\phi'_H} & L'_H, \end{array}$$

commutes, where the top arrow is defined as

$$\alpha^* - p^{-1}\pi(T_p)\beta^* + p^{-1}\gamma^*.$$

Extending scalars to $\mathbb{C}$, this map sends ω_g to $\omega_{g'}$ and $\bar{\omega}_{g^c}$ to $\bar{\omega}_{g'^c}$ where g' is defined by (4.4.3). In the case $\Sigma' = \Sigma \cup \{\ell\}$, the same assertion holds if $\mathcal{O}_0$ is chosen so that it contains $\pi(\alpha_\ell) = a_\ell(g')$ and the top arrow is defined as

$$\alpha^* - \pi(\alpha_\ell)^{-1}\beta^*.$$

We conclude that if $\{x, y\}$ is a basis for $H^1(X, \mathcal{O}_0)[\mathcal{P}_0]$ and the matrices A and A' are defined using the bases $\{x, y\}$ for $H^1(X, \mathcal{O}_0)[\mathcal{P}_0]$ and $\{\phi'_H x, \phi'_H y\}$

for $H^1(X', \mathcal{O}_0)[\mathcal{P}_0']$, then $A' = A$. Theorem 3.36 is now immediate from corollary 4.21. In fact, we have proved that

$$\eta_{\Sigma'} = \pi(\prod_{p \in \Sigma' - \Sigma} c_p)\eta_\Sigma.$$

□

Remark 4.25 Note that to obtain the inclusion stated in theorem 3.36 it suffices to apply theorem 4.18 for Σ, rather than both Σ and Σ' (cf. remark 4.22).

Wiles' argument: The method of [W3] sec. 2.2, like that of Ribet in [R4], is more direct than the one given above but does not explicitly illustrate the relation with values of L-functions. The approach is simply to compute the composite $\phi_T \phi_T'$. It suffices to consider the case $\Sigma' = \Sigma \cup \{p\}$ and we suppose first that $p \neq \ell$. Let Δ denote the endomorphism of J^3 defined by the matrix

$$\Delta = \begin{pmatrix} \alpha_* \\ \beta_* \\ \gamma_* \end{pmatrix} w_*'(\alpha^*, \beta^*, \gamma^*) w_*.$$

Using the relations $\alpha w' = w\gamma$ and $\beta w' = w\beta$, we find that

$$\Delta = \begin{pmatrix} \alpha_*\gamma^* & \alpha_*\beta^* & \alpha_*\alpha^* \\ \beta_*\gamma^* & \beta_*\beta^* & \beta_*\alpha^* \\ \gamma_*\gamma^* & \gamma_*\beta^* & \gamma_*\alpha^* \end{pmatrix},$$

which can be computed for example by considering its effect on the cotangent space $S_2(\Gamma_0(N))^3$. The result is that

$$\Delta = \begin{pmatrix} T_p^2 - (p+1) & pT_p & p(p+1) \\ pT_p & p(p+1) & pT_p \\ p(p+1) & pT_p & T_p^2 - (p+1) \end{pmatrix},$$

which we note commutes with the action of $\mathbb{T}_{\mathbf{Z}}$ on J^3. One then checks that

$$\begin{aligned} \phi_T \phi_T' &= (1, -p^{-1}T_p, p^{-1}) \ (\mathcal{T}_\ell(\Delta) \otimes_{\mathbf{Z}_\ell} \mathcal{O})_{\mathfrak{m}} \ \begin{pmatrix} 1 \\ -p^{-1}T_p \\ p^{-1} \end{pmatrix} \\ &= -p^{-2}(p-1)(T_p^2 - (p+1)^2). \end{aligned}$$

The case of $p = \ell$ is similar but simpler. One uses

$$\Delta = \begin{pmatrix} \alpha_* \\ \beta_* \end{pmatrix} w_*'(\alpha^*, \beta^*) w_* = \begin{pmatrix} T_\ell & \ell + 1 \\ \ell + 1 & T_\ell \end{pmatrix}$$

and gets

$$\phi_T \phi_T' = -(\alpha_\ell - \ell\alpha_\ell^{-1})(1 - \alpha_\ell^{-2}).$$

Thus in either case we find that $\phi_T \phi_T'$ acts on $L_T'[\mathcal{P}]$ by an element of $\mathcal{O} = \mathbb{T}_{\mathfrak{m}}/\mathcal{P}$ which is a unit times $\pi(c_p)$. Theorem 3.36 then follows from corollary 4.19 and lemma 4.24.

4.5 Homological results

In this section we sketch the proofs of theorem 4.18 and lemma 4.24, but shall often refer the reader to ch. 2 of [W3] for more details.

Multiplicities: We first consider theorem 4.18, generalizing a result proved by Mazur in sections II.14 and II.15 of [Maz1]. Recall that $L_{\mathcal{T}}$ (resp. L_H) is defined as $(\mathcal{T}_\ell(J_0(N))\otimes_{\mathcal{O}})_{\mathfrak{m}}$ (resp. $H^1(X_0(N),\mathcal{O})_{\mathfrak{m}}$) where $\mathfrak{m}$ is a maximal of the ideal of the Hecke algebra $\mathbb{T}_{\mathcal{O}}$ generated by the Hecke operators on $S_2(\Gamma_0(N))$. We are assuming moreover that $\rho_{\mathfrak{m}}$ is irreducible and that one of the following holds:

- ℓ does not divide N;
- ℓ^2 does not divide N and $T_\ell \notin \mathfrak{m}$.

We wish to prove that $L_{\mathcal{T}}$ and L_H are free over $\mathbb{T}_{\mathfrak{m}}$.

Since L_H and $L_{\mathcal{T}}$ are isomorphic as $\mathbb{T}_{\mathfrak{m}}$-modules, it suffices to prove that $L_{\mathcal{T}}$ is free. Furthermore, to prove that $L_{\mathcal{T}}$ is free, it suffices to prove that the localization of $\mathcal{T}_\ell(J_0(N))$ at $\mathfrak{m}\cap\mathbb{T}_{\mathbf{Z}_\ell}$ is free; i.e., we may replace $\mathcal{O}$ by $\mathbb{Z}_\ell$ and $\mathfrak{m}$ by its intersection with $\mathbb{T}_{\mathbf{Z}_\ell}$ and consider

$$\mathcal{T}_\ell(J_0(N))_{\mathfrak{m}}.$$

Since $\mathcal{T}_\ell(J_0(N))\otimes_{\mathbf{Z}_\ell}\mathbb{Q}_\ell$ is free is free of rank two over $\mathbb{T}_{\mathbf{Q}_\ell}$, it suffices to prove that

$$\dim_{\mathbb{T}/\mathfrak{m}}(L_{\mathcal{T}}/\mathfrak{m}L_{\mathcal{T}}) = 2.$$

Since

$$\mathcal{T}_\ell(J_0(N))/\ell\mathcal{T}_\ell(J_0(N)) \cong J_0(N)[\ell] \cong \mathrm{Hom}_{\mathbf{F}_\ell}(J_0(N)[\ell],\mathbb{F}_\ell)$$

as $\mathbb{T}_{\mathbf{F}_\ell}$-modules, it suffices to prove the following cases of [W3] thm. 2.1.

Theorem 4.26 *Let $\mathfrak{m}$ be a maximal ideal of $\mathbb{T}_{\mathbf{Z}_\ell}$. Suppose that $\rho_{\mathfrak{m}}$ is irreducible and that either*

(a) *ℓ does not divide N, or*

(b) *$\ell^2 \nmid N$ and $T_\ell \notin \mathfrak{m}$.*

Then

$$\dim_{\mathbb{T}_{\mathbf{Z}_\ell}/\mathfrak{m}} J_0(N)[\mathfrak{m}] = \dim_{\mathbb{T}_{\mathbf{Z}_\ell}/\mathfrak{m}}(J_0(N)[\ell]/\mathfrak{m}) = 2.$$

The proof of theorem 4.26 in case (b) requires more of the theory of group schemes and Néron models than we wish to delve into here. We shall therefore only explain the proof in the case (a) (which is [R5] thm. 5.2(b), see Ribet's paper for more details) and refer to [W3] sec. 2.1 for the general case. We remark however that the proof of theorem 3.42 appeals to theorem 3.36 only in the case $\Sigma = \emptyset$ (cf. remarks 4.22 and remark 4.25). Recall that ℓ

divides $N_\emptyset$ only if $\bar{\rho}$ is not good. So for the purposes of proving the Shimura-Taniyama conjecture for semistable elliptic curves, (b) can be replaced by the stronger hypothesis

(b') $\ell^2 \nmid N$ and $\rho_{\mathfrak{m}}|_{G_\ell}$ is not good.

The result in this case is due to Mazur and Ribet (the main result of [MRi]) and the proof is slightly easier than in the case of (b).

Returning to case (a), we appeal to a general property of the functor $\mathbb{D}$ of theorem 2.31 which follows from cor. 5.11 of [Oda].

Theorem 4.27 *Suppose that A is an abelian variety over $\mathbb{Z}_\ell$ (i.e., the Néron model of an abelian variety over $\mathbb{Q}_\ell$ with good reduction). There is a canonical and functorial isomorphism of vector spaces over $\mathbb{F}_\ell$*

$$\mathbb{D}(A(\bar{\mathbb{Q}}_\ell)[\ell]^*)^0 \cong \mathrm{Cot}_0(A/\mathbb{F}_\ell),$$

where Cot_0 denotes the cotangent space at the origin.

In the case that A is the Jacobian of a smooth proper curve X over $\mathbb{Z}_\ell$, we have also a canonical isomorphism

$$\mathrm{Cot}_0(A/\mathbb{F}_\ell) \cong H^0(X, \Omega^1_{X/\mathbb{F}_\ell})$$

and in the case $X = X_0(N)$ with $\ell \nmid N$, this provides an isomorphism of $\mathbb{T}_{\mathbb{F}_\ell}$-modules

$$\mathbb{D}(J_0(N)(\bar{\mathbb{Q}}_\ell)[\ell]^*)^0 \cong S_2(\Gamma_0(N), \mathbb{F}_\ell). \tag{4.5.1}$$

Consider now the action of $G_{\mathbb{Q}}$ on the points of

$$V = J_0(N)[\ell]^*[\mathfrak{m}] \cong (J_0(N)[\ell]/\mathfrak{m})^*.$$

By the argument in [Maz1] prop. II.14.2 or by the main result of [BLR], we know that every Jordan-Hölder constituent of the representation of $G_{\mathbb{Q}}$ on this $\mathbb{T}/\mathfrak{m}$-vector space is isomorphic to $\rho^*_{\mathfrak{m}} \cong \rho_{\mathfrak{m}}$. It follows that

$$\dim_{\mathbb{T}/\mathfrak{m}} \mathbb{D}(V) = 2 \dim_{\mathbb{T}/\mathfrak{m}} \mathbb{D}(V)^0$$

where we now regard V as a good G_ℓ-module. On the other hand (4.5.1) implies that

$$\mathbb{D}(V)^0 \cong S_2(\Gamma_0(N), \mathbb{F}_\ell)[\mathfrak{m}],$$

which is one-dimensional over $\mathbb{T}/\mathfrak{m}$ by the q-expansion principle (lemma 1.34 for example). We have now shown that $\mathbb{D}(V)$ is two-dimensional over $\mathbb{T}/\mathfrak{m}$, and it follows that so is V.

Ihara's lemma: We now sketch the proof of lemma 4.24, which proceeds by analyzing the behavior of the homology of modular curves under certain degeneracy maps. Note that it suffices to prove the lemma for $\phi_{\mathcal{T}}$.

If N and M are positive integers, then we let $\Gamma_1(N,M) = \Gamma_1(N) \cap \Gamma_0(M)$ and write $Y_1(N,M)$ (resp. $X_1(N,M)$) for the associated non-compactified (resp. compactified) modular curve. The key intermediate result is the following:

Lemma 4.28 *Suppose that N is a positive integer and p is a prime not dividing N.*

(a) *The map*

$$\begin{array}{ccc} H_1(X_1(N,p),\mathbb{Z}) & \to & H_1(X_1(N),\mathbb{Z})^2 \\ x & \mapsto & (\alpha_* x, \beta_* x) \end{array}$$

is surjective, where α is defined by $\tau \mapsto \tau$ and β by $\tau \mapsto p\tau$.

(b) *If $N > 3$ and ℓ is an odd prime different from p, then the sequence*

$$\begin{array}{ccccc} H_1(Y_1(Np,p^2),\mathbb{Z}_\ell) & \to & H_1(Y_1(Np),\mathbb{Z}_\ell)^2 & \to & H_1(Y_1(N),\mathbb{Z}_\ell) \\ x & \mapsto & (\alpha_{1*}x, \beta_{1*}x); & (y,z) \mapsto & \beta_{2*}y - \alpha_{2*}z \end{array}$$

is exact, where the maps α_1 and α_2 are defined by $\tau \mapsto \tau$ and β_1 and β_2 by $\tau \mapsto p\tau$.

To prove the lemma, one first translates the statement into one about the homology (or cohomology) of the congruence subgroups involved. Part (a) of the lemma is due to Ihara [Ih], but see also the proof of thm. 4.1 of [R4]. Part (b) due to Wiles is more elementary and is established in the course of proving lemma 2.5 of [W3] (see the sequence (2.13)). The result stated there is in terms of cohomology rather than homology, but the one given here is immediate from it.

Remark 4.29 A method of Khare [Kh], sec. 2 using modular symbols yields an alternate proof of part (b) (after localization at a non-Eisenstein maximal ideal).

To deduce lemma 4.24 from lemma 4.28 we also need the following result:

Lemma 4.30 *Suppose that $\mathcal{O}$ is the ring of integers of a finite extension of $\mathbb{Q}_\ell$. Let M be a positive integer and H a subgroup of $(\mathbb{Z}/M\mathbb{Z})^\times$. Let $\tilde{\mathbb{T}}_\mathcal{O} = \tilde{\mathbb{T}}_\mathbf{Z} \otimes \mathcal{O}$ where $\tilde{\mathbb{T}}_\mathbf{Z}$ is the subring of endomorphisms of $M_2(\Gamma_1(M))$ generated by the operators $\langle d \rangle$ and T_r for primes $r \nmid M$. If $\mathfrak{n}$ is a non-Eisenstein maximal ideal then the localization at $\mathfrak{n}$ of the natural map*

(a) *$H_1(Y_1(M),\mathcal{O}) \to H_1(X_1(M),\mathcal{O})$ is an isomorphism;*

(b) *$H_1(X_1(M),\mathcal{O}) \to H_1(X_H(M),\mathcal{O})$ is surjective.*

To prove part (a), one checks that

$$H_1(Y_1(M),\mathbb{Z}) \to H_1(X_1(M),\mathbb{Z})$$

is surjective and that $T_r = r + 1$ on the kernel for primes $r \equiv 1 \bmod M$ (cf. proposition 4.13, part (c)).

Part (b) is most easily proved by showing that $G_{\mathbb{Q}}$ acts trivially on the cokernel of the natural map

$$\mathcal{T}_\ell(J_1(M)) \to \mathcal{T}_\ell(J_H(M))$$

(using [LO], prop. 6 for example) and then applying lemma 4.12.

Finally we explain how to use lemmas 4.28 and 4.30 to prove lemma 4.24. It suffices to consider the case $\Sigma' = \Sigma \cup \{p\}$. We discuss only the more difficult case of $p \neq \ell$. We define $N = N_\Sigma$, $N' = Np^2$, $\mathfrak{m}$ and $\mathfrak{m}'$ as in the preceding section. Thus $\mathfrak{m}'$ is a maximal ideal of $\mathbb{T}'_{\mathcal{O}}$ where $\mathbb{T}'_{\mathbb{Z}}$ acts on $S_2(\Gamma_0(N'))$. All the maps we consider in the argument below will respect the action of the Hecke operators in $\tilde{\mathbb{T}}''_{\mathbb{Z}}$ where $\tilde{\mathbb{T}}''_{\mathbb{Z}}$ is the subring of endomorphisms of $M_2(\Gamma_1(Np, p^2))$ generated by the operators $\langle d \rangle$ and T_r for primes r not dividing Np. We let $\mathfrak{n}$ denote the preimage of $\mathfrak{m}'$ in $\tilde{\mathbb{T}}''_{\mathcal{O}} = \tilde{\mathbb{T}}''_{\mathbb{Z}} \otimes \mathcal{O}$. We first apply lemma 4.28(b) to show that

$$H_1(Y_1(Np, p^2), \mathcal{O})_{\mathfrak{n}} \to H_1(Y_1(Np), \mathcal{O})^2_{\mathfrak{n}} \to H_1(Y_1(N), \mathcal{O})_{\mathfrak{n}}$$

is exact. Next we apply lemma 4.30(a) to obtain the exactness of

$$H_1(X_1(Np, p^2), \mathcal{O})_{\mathfrak{n}} \to H_1(X_1(Np), \mathcal{O})^2_{\mathfrak{n}} \to H_1(X_1(N), \mathcal{O})_{\mathfrak{n}}.$$

Applying lemma 4.30(b) with $M = Np$, we find that

$$H_1(X_1(N, p^2), \mathcal{O})_{\mathfrak{n}} \to H_1(X_1(N, p), \mathcal{O})^2_{\mathfrak{n}} \to H_1(X_1(N), \mathcal{O})_{\mathfrak{n}}$$

is exact. We then apply lemma 4.28(a) to conclude that the map

$$\begin{array}{ccc} H_1(X_1(N, p^2), \mathcal{O})_{\mathfrak{n}} & \to & H_1(X_1(N), \mathcal{O})^3_{\mathfrak{n}} \\ x & \mapsto & (\alpha_* x, \beta_* x, \gamma_* x) \end{array}$$

is surjective where α, β and γ are defined respectively by $\tau \mapsto \tau$, $\tau \mapsto p\tau$ and $\tau \mapsto p^2\tau$. Finally, we use lemma 4.30(b) again (now with $M = N$) to get the surjectivity of

$$H_1(X_0(Np^2), \mathcal{O})_{\mathfrak{n}} \to H_1(X_0(N), \mathcal{O})^3_{\mathfrak{n}},$$

hence that of (4.4.4) upon localizing at $\mathfrak{m}'$.

5 Commutative algebra

In this section we collect some basic facts of commutative algebra that are used in the proof. We recall that $\mathcal{O}$ is the ring of integers in a finite extension K of $\mathbb{Q}_\ell$, and that $\mathcal{O}$ has residue field k. Let $\mathcal{C}_{\mathcal{O}}$ denote as in section 2.6 the category of complete noetherian local $\mathcal{O}$-algebras with residue field k.

5.1 Wiles' numerical criterion

In this section we state a numerical criterion discovered by Wiles for a map between two rings in $\mathcal{C}_{\mathcal{O}}$ to be an isomorphism.

Complete intersections: We say that a ring A in $\mathcal{C}_{\mathcal{O}}$ is *finite flat* if it is finitely generated and torsion free as an $\mathcal{O}$-module. A key ingredient in Wiles' isomorphism criterion is played by the concept of *complete intersection*, for which we give the following naive definition.

Definition 5.1 An object A in $\mathcal{C}_{\mathcal{O}}$ which is finite flat is called a *complete intersection* if it can be expressed as a quotient

$$A \simeq \mathcal{O}[[X_1, \ldots, X_n]]/(f_1, \ldots, f_n),$$

where there are as many relations as there are variables.

Remark 5.2 It is also true that if an object $\mathcal{O}[[Y_1, \ldots, Y_r]]/J$ in $\mathcal{C}_{\mathcal{O}}$ which is finite flat is a complete intersection, then necessarily J can be generated by r elements. See for example [Mat], thm. 21.2 (and lemma 5.11 below). We will not use this fact in this chapter, although a special case is proved in the course of establishing lemma 5.30.

The category $\mathcal{C}_{\mathcal{O}}^{\bullet}$: The numerical criterion of Wiles is stated more naturally in terms of rings A in the category $\mathcal{C}_{\mathcal{O}}$ which are endowed with some *extra structure*: namely, a surjective $\mathcal{O}$-algebra homomorphism $\pi_A : A \longrightarrow \mathcal{O}$. Let $\mathcal{C}_{\mathcal{O}}^{\bullet}$ be the category whose objects are pairs (A, π_A), where A is an object of $\mathcal{C}_{\mathcal{O}}$ and $\pi_A : A \longrightarrow \mathcal{O}$ is a surjective $\mathcal{O}$-algebra homomorphism, also called the *augmentation map* attached to A. Morphisms in $\mathcal{C}_{\mathcal{O}}^{\bullet}$ are local ring homomorphisms which are compatible in the obvious way with the augmentation maps. By abuse of notation one often omits mentioning the augmentation map π_A when talking of objects in $\mathcal{C}_{\mathcal{O}}^{\bullet}$, and simply uses A to denote (A, π_A), when this causes no confusion. Objects of $\mathcal{C}_{\mathcal{O}}^{\bullet}$ will also be called *augmented rings*.

The invariants Φ_A and η_A: One associates to an augmented ring (A, π_A) two basic invariants:

$$\begin{aligned} \Phi_A &= (\ker \pi_A)/(\ker \pi_A)^2; \\ \eta_A &= \pi_A(\mathrm{Ann}_A \ker \pi_A). \end{aligned}$$

Here $\mathrm{Ann}_A(I)$ denotes the annihilator ideal of the ideal I in A.

The invariant Φ_A can be thought of as a tangent space for the object A. (More precisely, it is the cotangent space of the scheme $\mathrm{spec}(A)$ at the point $\ker \pi_A$.) It is a finitely generated $\mathcal{O}$-module.

The invariant η_A seems less familiar at first sight. It is called the *congruence ideal*. (The reason for this terminology should become clearer shortly.) We are now ready to state Wiles' numerical criterion:

Theorem 5.3 *Let $\phi : R \longrightarrow T$ be a surjective morphism of augmented rings. Assume that T is finitely generated and torsion-free as an $\mathcal{O}$-module, and that $\eta_T \neq (0)$. (And hence, in particular, $\#(\mathcal{O}/\eta_T) < \infty$.) Then the following are equivalent:*

(a) *The inequality $\#\Phi_R \leq \#(\mathcal{O}/\eta_T)$ is satisfied.*

(b) *The equality $\#\Phi_R = \#(\mathcal{O}/\eta_T)$ is satisfied.*

(c) *The rings R and T are complete intersections, and the map $\phi : R \longrightarrow T$ is an isomorphism.*

Remark 5.4 The above theorem is slightly different from the one that appears in [W3], where it is assumed from the outset that the ring T is Gorenstein. In the form in which we state it above, the theorem is due to H. Lenstra [Len]. Our presentation follows Lenstra's very closely. For the original (and slightly different) point of view, the reader should consult the appendix of [W3].

Some examples: Before going further, it may be good to pause and consider some examples of objects of $\mathcal{C}_{\mathcal{O}}^{\bullet}$ and the invariants associated to them. While logically independent of the proof, the examples should help the reader develop some intuition. (For a systematic way to compute the tangent spaces Φ_A, see the paragraph at the end of section 5.2.)

Example 1:

$$\begin{aligned} A &= \{(a,b) \in \mathcal{O} \times \mathcal{O}, \quad a \equiv b \pmod{\lambda^n}\} \\ &\simeq \mathcal{O}[[T]]/(T(T-\lambda^n)), \qquad \text{with } \pi_A(a,b) := a. \\ \Phi_A &\simeq \mathcal{O}/\lambda^n\mathcal{O}, \qquad \eta_A = (\lambda^n). \end{aligned}$$

Example 2:

$$\begin{aligned} A &= \{(a,b,c) \in \mathcal{O} \times \mathcal{O} \times \mathcal{O}, a \equiv b \equiv c \pmod{\lambda}\} \\ &\simeq \mathcal{O}[[X,Y]]/(X(X-\lambda), Y(Y-\lambda), XY), \qquad \text{with } \pi_A(a,b,c) := a. \\ \Phi_A &\simeq \mathcal{O}/\lambda\mathcal{O} \times \mathcal{O}/\lambda\mathcal{O}, \qquad \eta_A = (\lambda). \end{aligned}$$

Example 3:

$$\begin{aligned} A &= \mathcal{O}[[X]]/(X^2), \qquad \text{with } \pi_A(f) := f(0). \\ \Phi_A &\simeq \mathcal{O}, \qquad \eta_A = 0. \end{aligned}$$

Example 4:

$$\begin{aligned} A &= \left\{(a,b,c,d) \in \mathcal{O} \times \cdots \times \mathcal{O}, \begin{array}{l} a \equiv b \equiv c \equiv d \pmod{\lambda}, \\ a+d \equiv b+c \pmod{\lambda^2} \end{array}\right\} \\ &\simeq \mathcal{O}[[X,Y]]/(X(X-\lambda), Y(Y-\lambda)), \qquad \text{with } \pi_A(a,b,c,d) := a. \end{aligned}$$

$$\Phi_A \simeq \mathcal{O}/\lambda\mathcal{O} \times \mathcal{O}/\lambda\mathcal{O}, \qquad \eta_A = (\lambda^2).$$

Example 5:

$$A = \mathcal{O}[[X_1, \ldots, X_n]], \qquad \text{with } \pi_A(f) := f(0).$$
$$\Phi_A \simeq \mathcal{O}^n, \qquad \eta_A = (0).$$

Example 6: $(\lambda = \ell \equiv -1 \pmod 4, \quad \mathcal{O} = \mathbb{Z}_\ell)$.

$$\begin{aligned} A &= \{(a, b + ci) \in \mathbb{Z}_\ell \times \mathbb{Z}_\ell[i], \quad a \equiv b \pmod{\ell^2}, \quad c \equiv 0 \pmod \ell\} \\ &\simeq \mathbb{Z}_\ell[[X]]/(X(X^2 + \ell^2)), \qquad \text{with } \pi_A(a, b + ci) := a. \end{aligned}$$
$$\Phi_A \simeq \mathbb{Z}/\ell^2\mathbb{Z}, \qquad \eta_A = (\ell^2).$$

Example 7:

$$A = \mathcal{O}[[T]]/(\lambda T) \simeq \mathcal{O} \oplus kT \oplus kT^2 \oplus \cdots, \qquad \text{with } \pi_A(f) = f(0).$$
$$\Phi_A \simeq k, \qquad \eta_A = (\lambda).$$

5.2 Basic properties of Φ_A and η_A

In this section we collect some of the basic properties of the invariants Φ_A and η_A, and prove the equivalence of (a) and (b) in theorem 5.3.

Behaviour of Φ_A under morphisms: The assignment $A \mapsto \Phi_A$ is a functor from the category $\mathcal{C}_\mathcal{O}^\bullet$ to the category of $\mathcal{O}$-modules; a morphism $A \longrightarrow B$ in $\mathcal{C}_\mathcal{O}^\bullet$ induces a homomorphism $\Phi_A \longrightarrow \Phi_B$ of $\mathcal{O}$-modules. Moreover, if $A \longrightarrow B$ is surjective, then so is the induced map on the tangent spaces. Therefore, when A maps surjectively onto B we have

$$\#\Phi_A \geq \#\Phi_B. \tag{5.2.1}$$

There is also a converse to this, which will be useful later:

Lemma 5.5 *If the homomorphism $\Phi_A \longrightarrow \Phi_B$ is surjective, then $A \longrightarrow B$ is also surjective.*

Proof: This follows from Nakayama's lemma. (Cf. [Ha], ch. II, sec. 7.4 and [Mat], th. 8.4.) □

Behaviour of η_A under (surjective) morphisms: Unlike the assignment $A \mapsto \Phi_A$, the assignment $A \longrightarrow \eta_A$ is not functorial, but it does have a nice behaviour under surjective morphisms: namely, if $\phi : A \longrightarrow B$ is surjective, then

$$\eta_A \subset \eta_B, \quad \text{i.e.,} \quad \#(\mathcal{O}/\eta_A) \geq \#(\mathcal{O}/\eta_B). \tag{5.2.2}$$

This is simply because in that case ϕ induces a map

$$\mathrm{Ann}_A \ker \pi_A \longrightarrow \mathrm{Ann}_B \ker \pi_B.$$

Relation between the invariants Φ_A and η_A: In general, we have the following inequality:

$$\#\Phi_A \geq \#(\mathcal{O}/\eta_A). \tag{5.2.3}$$

The key behind proving this identity is to interpret $\#\Phi_A$ in terms of Fitting ideals.

Digression on Fitting ideals: If R is a ring (in $\mathcal{C}_\mathcal{O}$, say) and M is a finitely generated R-module, we express M as a quotient of R^n for some n:

$$0 \longrightarrow M' \longrightarrow R^n \longrightarrow M \longrightarrow 0. \tag{5.2.4}$$

The *Fitting ideal* of M, denoted $\mathrm{Fitt}_R(M)$, is the ideal of R generated by the determinants $\det(v_1, \ldots, v_n)$, where the vectors $v_i \in R^n$ range over all possible choices of elements of $M' \subset R^n$. One checks that this ideal does not depend on the choice of exact sequence (5.2.4), and hence is an invariant of the R-module M. For example, if M is a finitely generated $\mathcal{O}$-module, we may write

$$M = \mathcal{O}^r \oplus \mathcal{O}/(\lambda^{n_1}) \oplus \mathcal{O}/(\lambda^{n_2}) \oplus \cdots \oplus \mathcal{O}/(\lambda^{n_k}),$$

with $n_1 \geq n_2 \geq \cdots \geq n_k$, and the sequence n_i is completely determined by M. The Fitting ideal $\mathrm{Fitt}_\mathcal{O}(M)$ is then the ideal of $\mathcal{O}$ generated by $\lambda^{n_1+\cdots+n_k}$, if $r = 0$, and is the 0-ideal if $r > 0$. Note in particular that, if M is a finite $\mathcal{O}$-module, then

$$\#M = \#(\mathcal{O}/\mathrm{Fitt}_\mathcal{O}(M)). \tag{5.2.5}$$

Furthermore, if M is any R-module, it follows directly from the definition that

$$\mathrm{Fitt}_R(M) \subset \mathrm{Ann}_R(M). \tag{5.2.6}$$

Fitting ideals behave well under tensor products: in particular, if M is a finitely generated A-module, where A is an object in $\mathcal{C}_\mathcal{O}^\bullet$, then:

$$\pi_A(\mathrm{Fitt}_A(M)) = \mathrm{Fitt}_\mathcal{O}(M \otimes_A \mathcal{O}), \tag{5.2.7}$$

where the tensor product is taken with respect to the augmentation map π_A. For more details and references on the Fitting ideal, see [Len] for example.

Now we are ready to prove equation (5.2.3). For, noting that $\Phi_A = \ker \pi_A \otimes_A \mathcal{O}$, where the tensor product is taken with respect to π_A, and applying equation (5.2.7) with $M = \ker \pi_A$, we have:

$$\mathrm{Fitt}_\mathcal{O}(\Phi_A) = \pi_A(\mathrm{Fitt}_A(\ker \pi_A)) \subset \pi_A(\mathrm{Ann}_A \ker \pi_A) = \eta_A, \tag{5.2.8}$$

where the containment follows from equation (5.2.6). Now the inequality (5.2.3) follows by combining (5.2.5) and (5.2.8).

As a consequence, we have

Corollary 5.6 *The statements* (*a*) *and* (*b*) *in theorem 5.3 are equivalent.*

Proof: If $R \longrightarrow T$ is a surjective map of augmented rings, then $\#\Phi_R \geq \#\Phi_T$, by equation (5.2.1). But equation (5.2.3) gives the inequality $\#\Phi_T \geq \#(\mathcal{O}/\eta_T)$. Hence the inequality $\#\Phi_R \geq \#(\mathcal{O}/\eta_T)$ always holds, so that (*a*) implies (*b*) in theorem 5.3. The reverse implication is clear. □

Remark 5.7 (Computing the tangent spaces Φ_A): Any object (A, π_A) in $\mathcal{C}_{\mathcal{O}}^{\bullet}$ can be expressed as a quotient of the object $U = \mathcal{O}[[X_1, \ldots, X_n]]$ of example 5 with augmentation map given by $\pi_U(f) = f(0)$. Indeed, one can take $a_1, \ldots, a_n$ to be A-module generators of the finitely generated A-module $\ker \pi_A$, and obtain the desired quotient map by sending X_i to a_i.

The tangent space Φ_U of U is a free $\mathcal{O}$-module of rank n which can be written down canonically as

$$\mathcal{O}X_1 \oplus \mathcal{O}X_2 \oplus \cdots \oplus \mathcal{O}X_n,$$

the natural map from $(\ker \pi_U)$ being simply the map which sends a power series $f \in U$ with no constant term to its degree 1 term, which we will denote by $\bar{f}$.

If A is expressed as a quotient $U/(f_1, \ldots, f_r)$, then one has

$$\Phi_A = \Phi_U/(\bar{f}_1, \ldots, \bar{f}_r). \tag{5.2.9}$$

5.3 Complete intersections and the Gorenstein condition

In this section we show that complete intersections satisfy a Gorenstein condition, and that (*c*) implies (*a*) and (*b*) in the statement of Wiles isomorphism criterion (theorem 5.3).

Definition 5.8 An object A in $\mathcal{C}_{\mathcal{O}}$ which is finite flat is said to be *Gorenstein* if

$$\mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O}) \simeq A \quad \text{as } A\text{-modules.}$$

Proposition 5.9 *Suppose A in $\mathcal{C}_{\mathcal{O}}$ is finite flat. If A is a complete intersection, then A is Gorenstein.*

The remainder of this section will be devoted to proving proposition 5.9. Since the proof is a bit long and involved, and the concepts it uses are not used elsewhere, the reader is advised on a first reading to take it on faith

and skip to the next section. A more direct proof of proposition 5.9 which bypasses the arguments of this section is explained in [Len].

We let A be a ring which is finite flat, and is a complete intersection. (Hence, A can be written as $\mathcal{O}[[X_1, \dots, X_n]]/(f_1, \dots, f_n)$.) We assume that the augmentation map for A is induced from the map on $\mathcal{O}[[X_1, \dots, X_n]]$ sending f to $f(0)$. This implies that $f_i(0) = 0$, i.e., the f_i have no constant term.

We recall some definitions from commutative algebra that we will need. An ideal I of a local ring R is said to be *primary* if $I \neq R$ and every zero divisor in R/I is nilpotent. If $(x_1, \dots, x_n)$ generates a primary ideal of R, and $n = \dim R$, then $(x_1, \dots, x_n)$ is called a *system of parameters* for R.

Lemma 5.10 *The sequence $(f_1, \dots, f_n, \lambda)$ is a system of parameters for $U = \mathcal{O}[[X_1, \dots, X_n]]$.*

Proof: The quotient ring $U/(f_1, \dots, f_n, \lambda)$ is local and is finitely generated as a k-vector space; therefore every element in its maximal ideal is nilpotent. □

A sequence $(x_1, \dots, x_n)$ in a ring R is said to be a *regular sequence* if x_i is not a zero-divisor in $R/(x_1, \dots, x_{i-1})$ for $i = 1, \dots, n$.

Lemma 5.11 *The sequence $(f_1, \dots, f_n)$ is a regular sequence for U.*

Proof: The ring U is Cohen Macaulay, since $\lambda, X_1, \dots, X_n$ is a system of parameters of U which is also a regular sequence. Hence, by theorem 17.4 (iii) of [Mat], $(f_1, \dots, f_n, \lambda)$ is a regular sequence in U. A fortiori, the sequence $(f_1, \dots, f_n)$ is also a regular sequence. □

The proofs of lemma 5.10 and 5.11 use only the fact that A is finitely generated as an $\mathcal{O}$-module, and not that A is flat. As a corollary of this proof, we therefore have:

Corollary 5.12 *If R is an object of $\mathcal{C}_{\mathcal{O}}^{\bullet}$ which is finitely generated as an $\mathcal{O}$-module, and $R \simeq \mathcal{O}[[X_1, \dots, X_n]]/(f_1, \dots, f_n)$, then R is also flat, and hence is a complete intersection.*

To go further, we will introduce the Koszul complex

$$K(\underline{x}, R) := \oplus_{p=0}^{n} K_p(\underline{x}, R)$$

associated to a local ring R and a sequence $\underline{x} = (x_1, \dots, x_n)$ of elements in its maximal ideal. For more details on the Koszul complex and its relation to regular sequences, the reader can consult [Mat], especially §16, or [Bour] X, or [Se3]. This complex is defined to be the free graded differential algebra generated by symbols $u_1, \dots, u_n$:

$$K_p(\underline{x}, R) := \oplus_{i_1 < i_2 < \cdots < i_p} R \cdot u_{i_1} \wedge \cdots \wedge u_{i_p},$$

with differential $d : K_p \longrightarrow K_{p-1}$ defined by

$$d(u_{i_1} \wedge \cdots \wedge u_{i_p}) = \sum_{t=1}^{p} (-1)^t x_t \cdot u_{i_1} \wedge \cdots \wedge u_{i_{t-1}} \wedge u_{i_{t+1}} \wedge \cdots \wedge u_{i_p}.$$

We denote by $H_p(\underline{x}; R)$ the homology groups of this complex. We record here the main properties of this complex that we will use.

Proposition 5.13 (a) $H_0(\underline{x}; R) = R/(\underline{x})$.

(b) *There is a long exact homology sequence*

$$\cdots \longrightarrow H_p(\underline{x}; R) \longrightarrow H_p(\underline{x}, x_{n+1}; R) \longrightarrow H_{p-1}(\underline{x}; R) \xrightarrow{\pm x_{n+1}}$$

$$H_{p-1}(\underline{x}; R) \longrightarrow H_{p-1}(\underline{x}, x_{n+1}; R) \longrightarrow H_{p-2}(\underline{x}; R) \longrightarrow \cdots$$

(c) $H_p(\underline{x}; R)$ *is annihilated by the ideal* $(\underline{x})$, *i.e., it has a natural* $R/(\underline{x})$*-module structure.*

(d) *If* $\underline{x}$ *is a regular sequence, then* $H_p(\underline{x}; R) = 0$ *for all* $p > 0$ *(i.e., the complex* $K_p(\underline{x}; R)$ *is a free resolution of* $R/(\underline{x})$*.)*

Proof: The first assertion follows directly from the definition. For (*b*) and (*c*), see [Mat], th. 16.4. The assertion (*d*) can be proved by a direct induction argument on n, using the long exact homology sequence: For $p > 1$, this sequence becomes

$$0 \longrightarrow H_p(\underline{x}, x_{n+1}; R) \longrightarrow 0,$$

and for $p = 1$, it is

$$0 \longrightarrow H_1(\underline{x}, x_{n+1}; R) \longrightarrow H_0(\underline{x}; R) \xrightarrow{x_{n+1}} H_0(\underline{x}; R).$$

But the assumption that $\underline{x}, x_{n+1}$ is a regular sequence means that multiplication by x_{n+1} is injective on $H_0(\underline{x}; R) = R/(\underline{x})$. Hence, the assertion (*d*) follows. □

Now, we turn to the proof of proposition 5.9, following a method of Tate which is explained in the appendix of [MRo]. For any ring R, write $R[[\underline{X}]] := R[[X_1, \ldots, X_n]]$. Let $a_1, \ldots, a_n$ be the images in A of $X_1, \ldots, X_n$ by the natural map

$$\alpha : \mathcal{O}[[\underline{X}]] \longrightarrow A = \mathcal{O}[[\underline{X}]]/(f_1, \ldots, f_n),$$

and let

$$\beta : A[[\underline{X}]] \longrightarrow A$$

be the natural map which sends X_i to a_i. The sequence $(g_i) = (X_i - a_i)$ generates the kernel of β. Since the f_i, viewed as polynomials in $A[[\underline{X}]]$, also belong to $\ker \beta$, we have:

$$(f_1, \ldots, f_n) = (g_1, \ldots, g_n)M, \tag{5.3.1}$$

where M is an $n \times n$ matrix with coefficients in $A[[\underline{X}]]$. Let

$$D = \det(M) \in A[[\underline{X}]].$$

Our goal is to construct an A-module isomorphism

$$\mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O}) \longrightarrow A.$$

We begin by constructing a surjective ($\mathcal{O}$-linear) map

$$\mathrm{Hom}_{\mathcal{O}[[\underline{X}]]}(A[[\underline{X}]], \mathcal{O}[[\underline{X}]]) \longrightarrow A.$$

Lemma 5.14 *(Tate): The function $\Phi(f) = \alpha(f(D))$ induces an isomorphism of $\mathcal{O}[[\underline{X}]]$-modules*

$$\mathrm{Hom}_{\mathcal{O}[[\underline{X}]]}(A[[\underline{X}]], \mathcal{O}[[\underline{X}]])/(g_1, \ldots, g_n) \longrightarrow A.$$

Proof: By lemma 5.11, the sequence $(\underline{f}) = (f_1, \ldots, f_n)$ is a regular $\mathcal{O}[[\underline{X}]]$-sequence. One can see directly from the definition that the sequence $(\underline{g}) = (g_i) = (X_i - a_i)$ is a regular $A[[\underline{X}]]$-sequence. Let $K(\underline{f})$ and $K(\underline{g})$ be the Koszul complexes associated to these two sequences. It follows from proposition 5.13 that the Koszul complex $K(\underline{f})$ is a resolution of A by free $\mathcal{O}[[\underline{X}]]$-modules, and the Koszul complex $K(\underline{g})$ is a resolution of A by free $A[[\underline{X}]]$-modules, and hence a fortiori, by free $\mathcal{O}[[\underline{X}]]$-modules. We define a map $\Phi : K(\underline{f}) \longrightarrow K(\underline{g})$ of complexes by letting

$$\Phi_0 : K_0(\underline{f}) \longrightarrow K_0(\underline{g})$$

be the natural inclusion of $\mathcal{O}[[\underline{X}]]$ into $A[[\underline{X}]]$, and letting

$$\Phi_1 : K_1(\underline{f}) \longrightarrow K_1(\underline{g})$$

be the map defined by

$$(\Phi_1(u_1), \ldots, \Phi_1(u_n)) = (v_1, \ldots, v_n)M,$$

and extending it by skew-linearity to a map of exterior algebras. One can check that the resulting map Φ is a morphism of complexes which induces the identity map $A \longrightarrow A$, and satisfies

$$\Phi_n(u_1 \wedge \cdots \wedge u_n) = D \cdot v_1 \wedge \cdots \wedge v_n.$$

Applying the functor $\mathrm{Hom}_{\mathcal{O}[[\underline{X}]]}(-, \mathcal{O}[[\underline{X}]])$ to these two free resolutions, and taking the homology of the resulting complexes, we find that since Φ is a homotopy equivalence of complexes it induces an isomorphism on the cohomology, and in particular, on the nth cohomology:

$$\Phi_n : \mathrm{Hom}_{\mathcal{O}[[\underline{X}]]}(A[[\underline{X}]], \mathcal{O}[[\underline{X}]])/(g_1, \ldots, g_n) \xrightarrow{\simeq}$$

$$\mathrm{Hom}_{\mathcal{O}[[\underline{X}]]}(\mathcal{O}[[\underline{X}]], \mathcal{O}[[\underline{X}]])/(f_1, \ldots, f_n) \simeq A,$$

which is given explicitly by the formula:

$$\Phi_n(f) = \alpha(f(D)).$$

□

We finally come to the proof of proposition 5.9, which we can state in a more precise form.

Lemma 5.15 *The map* $\Psi : \mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O}) \longrightarrow A$ *defined by* $\Psi(f) = \alpha(\tilde{f}(D))$, *where* $\tilde{f} : A[[\underline{X}]] \longrightarrow \mathcal{O}[[\underline{X}]]$ *is the base change of* f, *is an* A*-module isomorphism, and hence,* A *is Gorenstein.*

Proof: The key point is to show that Ψ is A-linear. By definition, if

$$a = \alpha(a') \in A, \text{ with } a' \in \mathcal{O}[[\underline{X}]],$$

then

$$\Psi(af) = \alpha(\tilde{f}(aD)) = \alpha(\tilde{f}((a - a')D)) + \alpha(\tilde{f}(a'D)).$$

Since $a - a' \in \ker \beta$, it can be written as an $A[[\underline{X}]]$-linear combination of the g_i. By multiplying the relation

$$(f_1, \ldots, f_n) = (g_1, \ldots, g_n)M$$

by the matrix $D \cdot M^{-1} \in M_n(A[[\underline{X}]])$, one sees that the Dg_i can be written as $A[[\underline{X}]]$-linear combinations of the f_i's. Hence, so can the expression $(a-a')D$. By the $\mathcal{O}[[\underline{X}]]$-linearity of $\tilde{f}$, and the fact that each f_i belongs to $\ker \alpha$, it follows that

$$\alpha(\tilde{f}((a - a')D)) = 0.$$

Therefore,

$$\Psi(af) = \alpha(a'\tilde{f}(D)) = a\Psi(f).$$

This shows that Ψ is A-linear. To show that Ψ is surjective, observe that if $f_1, \ldots, f_r$ is a $\mathcal{O}$-basis of $\mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O})$, then $\tilde{f}_1, \ldots, \tilde{f}_r$ is a $\mathcal{O}[[\underline{X}]]$-basis for $\mathrm{Hom}_{\mathcal{O}[[\underline{X}]]}(A[[\underline{X}]], \mathcal{O}[[\underline{X}]])$. Hence, for all $a \in A$, there exist $p_1, \ldots, p_r$ such that

$$\Phi_n(p_1\tilde{f}_1 + \cdots + p_r\tilde{f}_r) = a.$$

But this means that

$$\Psi(\alpha(p_1)f_1 + \cdots + \alpha(p_r)f_r) = a,$$

so that Ψ is surjective. Finally, since $\mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O})$ and A are free $\mathcal{O}$-modules of the same finite rank, and Ψ is surjective, it must also be injective. Hence Ψ is an isomorphism, as was to be shown. □

Example 5.16 Let $A \subset \mathcal{O} \times \mathcal{O}$ be the ring of example 1 in section 5.1. Then, $f = T^2 - \lambda^n T$, and $g = T - (0, \lambda^n)$. Hence, $f = (T - (\lambda^n, 0))g$, so that $D = T - (\lambda^n, 0)$. It follows that

$$\Phi(h) = \alpha(\tilde{h}(T - (\lambda^n, 0))) = \alpha(Th(1,1) - h(\lambda^n, 0)) = (0, \lambda^n)h(1,1) - h(\lambda^n, 0).$$

The reader can check directly that Φ is indeed an A-linear isomorphism from the A-module $\mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O})$ to A.

Example 5.17 Let $A = \mathcal{O}[\epsilon]/(\epsilon^2)$ be the ring of example 3 in section 5.1. Then, $f = T^2$, and $g = T - \epsilon$. Hence, $f = (T + \epsilon)g$, so that $D = T + \epsilon$. It follows that

$$\Phi(f) = \alpha(\tilde{f}(T + \epsilon)) = \alpha(Tf(1) + f(\epsilon)) = \epsilon f(1) + f(\epsilon).$$

Exercise 5.18 Write down explicitly an isomorphism $A \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O})$ for the complete intersection $A = \mathcal{O}[[X, Y]]/(X(X - \lambda), Y(Y - \lambda))$ of example 4 of section 5.1.

5.4 The Congruence ideal for complete intersections

Let A be an object of $\mathcal{C}_{\mathcal{O}}^{\bullet}$, which is finite flat and is a complete intersection as in the previous section, so that $A \simeq \mathcal{O}[[X_1, \ldots, X_n]]/(f_1, \ldots, f_n)$.

Using the result of the previous section, we will give an explicit formula for computing η_A in this case, and prove that (c) implies (b) in theorem 5.3.

Let $A^\vee := \mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O})$, and let $\pi_A^\vee : \mathcal{O}^\vee \longrightarrow A^\vee$ be the dual map. From the Gorenstein property of A, we may identify $A^\vee$ with A (as A-modules). Fix any identification $\Psi : A^\vee \xrightarrow{\simeq} A$. (Any two such differ by a unit in A.) One checks that

$$\Psi \pi_A^\vee(\mathcal{O}^\vee) = \mathrm{Ann}_A \ker \pi_A.$$

Hence, η_A is the image of the map $\pi_A \Psi \pi_A^\vee$. By using the explicit construction of Ψ given in lemma 5.15, we find:

$$\pi_A \Psi \pi_A^\vee(\mathcal{O}^\vee) = \pi \alpha \tilde{\pi}(D),$$

where D is the determinant defined in section 5.3. By a direct calculation using equation (5.3.1), one sees that the right hand side is equal to $\det(\partial f_i / \partial X_j(0))$. Hence we have shown:

Proposition 5.19 *Suppose that A is an object of $\mathcal{C}_{\mathcal{O}}^{\bullet}$ which is finite flat. If $A = \mathcal{O}[[X_1, \ldots, X_n]]/(f_1, \ldots, f_n)$ is a complete intersection, then*

$$\eta_A = (\det(\partial f_i / \partial X_j(0))).$$

This proposition implies:

Corollary 5.20 *Let A in $\mathcal{C}_{\mathcal{O}}^{\bullet}$ be finite flat. If A is a complete intersection, then $\#\Phi_A = \#(\mathcal{O}/\eta_A)$. Hence, the statement (c) implies statement (b) in theorem 5.3.*

Proof: The equation (5.2.9) of remark 5.7 implies that

$$\#\Phi_A = \#\mathcal{O}/(\det(\partial f_i/\partial X_j(0))).$$

The result follows. □

5.5 Isomorphism theorems

The usefulness of the notion of complete intersections comes from the following two (vaguely stated) principles:
1. Isomorphisms to complete intersections can often be recognized by looking at their effects on the tangent spaces.
2. Isomorphisms from complete intersections can often be recognized by looking at their effects on the invariants η.

These vague principles are made precise in theorems 5.21 and 5.24 respectively.

Theorem 5.21 *Let $\phi : A \longrightarrow B$ be a surjective morphism of augmented rings, with B a (finite, flat) complete intersection. If ϕ induces an isomorphism from Φ_A to Φ_B, and these modules are finite, then ϕ is an isomorphism.*

Remark 5.22 Let $A = \mathcal{O}[[X,Y]]/(X(X-\lambda), Y(Y-\lambda))$ be the ring of example 4, let $B = \mathcal{O}[[X,Y]]/(X(X-\lambda), Y(Y-\lambda), XY)$ be the ring of example 2, and let $\phi : A \longrightarrow B$ be the natural projection. The map ϕ induces an isomorphism $\Phi_A \longrightarrow \Phi_B$, even though ϕ is not an isomorphism. The assumption that B is a complete intersection is crucial for concluding that ϕ is an isomorphism.

Remark 5.23 The natural map

$$\mathcal{O}[[X]]/(X^3) \longrightarrow \mathcal{O}[[X]]/(X^2)$$

is a surjective morphism inducing an isomorphism on tangent spaces, and the target ring is a complete intersection. Yet this map is *not* an isomorphism. This shows that the assumption on the finiteness of the tangent spaces cannot be dispensed with.

Proof of theorem 5.21: Recall that $U = \mathcal{O}[[X_1, \ldots, X_n]]$ is the augmented ring of example 5 of section 5.1. Let

$$\nu_B : U \longrightarrow B$$

be a surjective morphism of augmented rings with $\ker \nu_B = (f_1, \ldots, f_n)$. Let $b_1, \ldots, b_n \in \ker \pi_B$ denote the images of $X_1, \ldots, X_n$ by ν_B, and let $a_1, \ldots, a_n \in \ker \pi_A$ denote inverse images of $b_1, \ldots, b_n$ by ϕ. Since ϕ is an isomorphism on tangent spaces, the elements a_i generate $(\ker \pi_A)/(\ker \pi_A)^2$. Hence the morphism

$$\nu_A : \mathcal{O}[[X_1, \ldots, X_n]] \longrightarrow A$$

defined by $\nu_A(X_i) = a_i$ induces a surjection $\Phi_U \longrightarrow \Phi_A$, and so it is surjective, by lemma 5.5.

We claim that $\ker \nu_B$ is contained in $\ker \nu_A$ (and hence, $\ker \nu_B = \ker \nu_A$). For, let $g_1, \ldots, g_n$ be elements of $\ker \nu_A$ whose linear terms $\bar{g}_1, \ldots, \bar{g}_n$ generate the kernel of

$$\bar{\nu}_A : \Phi_U \longrightarrow \Phi_A.$$

Since $\ker \nu_A \subset \ker \nu_B$, it follows that there exists an $n \times n$ matrix $M \in M_n(U)$ with entries in U such that

$$(g_1, \ldots, g_n) = (f_1, \ldots, f_n)M.$$

Let $\bar{M}$ be the matrix of constant terms of the matrix M. Then we have

$$(\bar{g}_1, \ldots, \bar{g}_n) = (\bar{f}_1, \ldots, \bar{f}_n)\bar{M}.$$

Since $(\bar{g}_1, \ldots, \bar{g}_n)$ and $(\bar{f}_1, \ldots, \bar{f}_n)$ generate the same submodules of rank n and finite index in Φ_U, it follows that $\det \bar{M}$ is a unit in $\mathcal{O}$. Hence, M is invertible, and therefore the f_i can be expressed as a U-linear combination of the g_j. This implies that $\ker \nu_B \subset \ker \nu_A$. Now we see that $\nu_A \nu_B^{-1}$ gives a well-defined inverse to ϕ, so that ϕ is an isomorphism. □

Theorem 5.24 *Let $\phi : A \longrightarrow B$ be a surjective morphism of augmented rings. Suppose that A and B are finite flat, and that A a complete intersection. If $\eta_A = \eta_B \neq 0$, then ϕ is an isomorphism.*

Remark 5.25 The torsion-freeness assumption on B is essential: if n is large enough, then $B = A/(\ker \pi_A)^n$ satisfies $\eta_A = \eta_B$, although the natural map $A \longrightarrow B$ is not injective when $A \neq \mathcal{O}$.

Proof: By proposition 5.9, A is Gorenstein, i.e., it satisfies

$$A^\vee := \mathrm{Hom}_{\mathcal{O}}(A, \mathcal{O}) \simeq A \text{ as } A\text{-modules}.$$

Now, we observe that

$$\ker \pi_A \cap \mathrm{Ann}_A \ker \pi_A = 0, \tag{5.5.1}$$

and likewise for B. For, let x be a non-zero element of η_A, and let $x' \in \mathrm{Ann}_A \ker \pi_A$ satisfy $\pi_A(x') = x$. For all $a \in \ker \pi_A \cap \mathrm{Ann}_A \ker \pi_A$, we have

$$0 = a(x - x') = ax,$$

the first equality because a belongs to $\mathrm{Ann}_A \ker \pi_A$ and $(x - x')$ belongs to $\ker \pi_A$, the second equality because a belongs to $\ker \pi_A$ and x' belongs to $\mathrm{Ann}_A \ker \pi_A$. Hence a belongs to the $\mathcal{O}$-torsion submodule of A, and therefore is 0.

It follows from (5.5.1) that the homomorphism π_A (resp. π_B) induces an isomorphism from $\mathrm{Ann}_A \ker \pi_A$ (resp. $\mathrm{Ann}_B \ker \pi_B$) to η_A (resp. η_B). Since η_A is isomorphic to η_B, it follows that ϕ induces an isomorphism from $\mathrm{Ann}_A \ker \pi_A$ to $\mathrm{Ann}_B \ker \pi_B$, i.e.,

$$\phi \mathrm{Ann}_A \ker \pi_A = \mathrm{Ann}_B \ker \pi_B.$$

From (5.5.1) it also follows a fortiori that

$$\ker \phi \cap \mathrm{Ann}_A \ker \pi_A = 0,$$

hence there is an exact sequence of A-modules:

$$0 \longrightarrow \ker \phi \oplus \mathrm{Ann}_A \ker \pi_A \longrightarrow A. \tag{5.5.2}$$

The cokernel of the last map is

$$A/(\ker \phi \oplus \mathrm{Ann}_A \ker \pi_A) \simeq B/(\phi \mathrm{Ann}_A \ker \pi_A) \simeq B/(\mathrm{Ann}_B \ker \pi_B),$$

which is torsion-free, since there is a natural injection

$$B/(\mathrm{Ann}_B \ker \pi_B) \hookrightarrow \mathrm{End}_{\mathcal{O}}(\ker \pi_B).$$

Hence, the exact sequence (5.5.2) splits over $\mathcal{O}$. Taking $\mathcal{O}$ duals in (5.5.2) and using the Gorenstein condition for A, we thus get an exact sequence of A-modules:

$$A \longrightarrow (\ker \phi)^\vee \oplus (\mathrm{Ann}_A \ker \pi_A)^\vee \longrightarrow 0.$$

Applying the functor $- \otimes_A k$ (relative to the map $A \longrightarrow k$), we find

$$1 = \dim_k(A \otimes_A k) \geq \dim_k((\ker \phi)^\vee \otimes_A k) + \dim_k((\mathrm{Ann}_A \ker \pi_A)^\vee \otimes_A k).$$

Since $\eta_A \neq 0$, it follows that $(\mathrm{Ann}_A \ker \pi_A)^\vee \otimes_A k \neq 0$, and hence we must have

$$(\ker \phi)^\vee \otimes_A k = 0.$$

Therefore by Nakayama's lemma and duality, $\ker \phi = 0$, which proves the theorem. □

5.6 A resolution lemma

It turns out that objects in $\mathcal{C}_{\mathcal{O}}^{\bullet}$ can be "resolved" (in a weak sense) by a complete intersection, namely,

Theorem 5.26 *Let A be an augmented ring which is finite flat over $\mathcal{O}$. Then there is a morphism $\tilde{A} \longrightarrow A$ in the category $\mathcal{C}_{\mathcal{O}}^{\bullet}$ such that:*

(a) *the ring $\tilde{A}$ is finite flat over $\mathcal{O}$ and is a complete intersection;*

(b) *the map $\tilde{A} \longrightarrow A$ induces an isomorphism $\Phi_{\tilde{A}} \longrightarrow \Phi_A$.*

Proof: Write A as a quotient of $U = \mathcal{O}[[X_1, \ldots, X_n]]$ (with $\pi_U : U \longrightarrow \mathcal{O}$ the map which sends each X_i to 0.) Let $f_1, \ldots, f_n$ be elements in the kernel of the natural map $U \longrightarrow A$, such that $\bar{f}_1, \ldots, \bar{f}_n$ generate the kernel of $\Phi_U \longrightarrow \Phi_A$. Now letting

$$\tilde{A} = U/(f_1, \ldots, f_n)$$

would give the desired ring $\tilde{A}$, provided $\tilde{A}$ is finitely generated: indeed, the flatness of $\tilde{A}$ would follow from corollary 5.12.

Thus we need to show that the f_i can be chosen so that $\tilde{A}$ is finitely generated. Let $a_1, \ldots, a_n$ be $\mathcal{O}$-module generators of the finitely generated module $\ker \pi_A$, and define a homomorphism ϕ from the polynomial ring

$$V = \mathcal{O}[X_1, \ldots, X_n]$$

to A by sending X_j to a_j. Clearly ϕ is surjective. Let $f_1, \ldots, f_n$ be elements of $\ker \phi$ chosen as above, and let m denote their maximal degree. Since the elements a_i^2 belong to $\ker \pi_A$, we may write

$$a_i^2 = h_i(a_1, \ldots, a_n),$$

where $h_i(X_1, \ldots, X_n)$ is a linear polynomial. Now, replacing the relations f_i by the relations

$$f_i + X_i^m h_i - X_i^{m+2},$$

we find that the ring $V/(f_1, \ldots, f_n)$ is a finitely generated $\mathcal{O}$-module: it can be generated by the images of the monomials of degree $\leq n(m+1)$, since the relations allow us to rewrite any monomial of higher degree in terms of ones of lower degree. Completing at the ideal $(\lambda, X_1, \ldots, X_n)$, we find that the ring

$$\tilde{A} = U/(f_1, \ldots, f_n)$$

has the desired properties: the natural homomorphism from $\tilde{A}$ to A induces an isomorphism on the tangent spaces, since the linear terms of the f_i generate the kernel of the induced map $\Phi_U \longrightarrow \Phi_A$ on the tangent spaces, and $\tilde{A}$ is a finitely generated $\mathcal{O}$-module, since $V/(f_1, \ldots, f_n)$ is. □

5.7 A criterion for complete intersections

The results we have accumulated so far allow us to give an important criterion for an object A to be a complete intersection:

Theorem 5.27 *Let A be an augmented ring which is a finitely generated torsion-free $\mathcal{O}$-module. If $\#\Phi_A \leq \#(\mathcal{O}/\eta_A) < \infty$, then A is a complete intersection.*

Proof: Let $\phi : \tilde{A} \longrightarrow A$ be the surjective morphism given by the resolution theorem (theorem 5.26). Then we have

$$\#(\mathcal{O}/\eta_A) \geq (\#\Phi_A) = (\#\Phi_{\tilde{A}}) \geq \#(\mathcal{O}/\eta_{\tilde{A}}),$$

where the first inequality is by assumption, the second by the choice of $\tilde{A}$, and the third is by the equation (5.2.3). On the other hand, by equation (5.2.2), we have

$$\#(\mathcal{O}/\eta_{\tilde{A}}) \geq \#(\mathcal{O}/\eta_A).$$

It follows that

$$\eta_A = \eta_{\tilde{A}},$$

so that ϕ is an isomorphism by theorem 5.24. It follows that A is a complete intersection. □

5.8 Proof of Wiles' numerical criterion

Theorem 5.28 *Let R and T be augmented rings such that T is a finitely generated torsion-free $\mathcal{O}$-module, and let $\phi : R \longrightarrow T$ be a surjective morphism. If*

$$\#\Phi_R \leq \#(\mathcal{O}/\eta_T) < \infty,$$

then R and T are complete intersections, and ϕ is an isomorphism.

Proof: We have:

$$\#(\mathcal{O}/\eta_T) \leq \#\Phi_T \leq \#\Phi_R \leq \#(\mathcal{O}/\eta_T),$$

where the first inequality is by equation (5.2.3), the second follows from the surjectivity of ϕ, and the third is by hypothesis. Therefore,

$$\#\Phi_T = \#(\mathcal{O}/\eta_T),$$

and hence T is a complete intersection by theorem 5.27. Since the orders of Φ_R and Φ_T are the same, ϕ induces an isomorphism between them. Hence ϕ is an isomorphism $R \longrightarrow T$, by theorem 5.21. This completes the proof. □

Theorem 5.28 shows that the statements (a) and (b) in theorem 5.3 imply the statement (c). Combining corollaries 5.6, 5.20, and theorem 5.28 completes the proof of theorem 5.3.

5.9 A reduction to characteristic ℓ

Let $\mathcal{C}_k$ be the category of complete local noetherian k-algebras with residue field k. Again all morphisms are assumed to be local. There is a natural functor $A \mapsto \bar{A}$ from $\mathcal{C}_\mathcal{O}$ to $\mathcal{C}_k$ which send A to $\bar{A} := A/\lambda$.

We say that an object A of $\mathcal{C}_k$ which is finite dimensional as a k-vector space is a complete intersection if it is isomorphic to a quotient

$$A = k[[X_1, \ldots, X_r]]/(f_1, \ldots, f_r).$$

Note that if an object A of $\mathcal{C}_{\mathcal{O}}$ is a complete intersection, then $\bar{A}$ is a complete intersection in $\mathcal{C}_k$. As a partial converse, we have:

Lemma 5.29 *Suppose that $R \mapsto T$ is a map in the category $\mathcal{C}_{\mathcal{O}}$, and that T is finitely generated and free as an $\mathcal{O}$-module. Then $R \longrightarrow T$ is an isomorphism of complete intersections, if and only if $\bar{R} \longrightarrow \bar{T}$ is.*

Proof: This is an exercise and is left to the reader. □

We now come to the proof of lemma 3.39 of section 3.4:

Lemma 5.30 *Suppose that $K \subset K'$ are local fields with rings of integers $\mathcal{O} \subset \mathcal{O}'$ and that A is an object of $\mathcal{C}_{\mathcal{O}}$ which is finitely generated and free as an $\mathcal{O}$-module. Then A is a complete intersection if and only if $A \otimes_{\mathcal{O}} \mathcal{O}'$ is.*

Proof: One implication is clear. Let k and k' be the residue fields of $\mathcal{O}$ and $\mathcal{O}'$ respectively. By lemma 5.29 it is enough to prove that, if R is an object of $\mathcal{C}_k$ which is finite dimensional as a k-vector space, then

$$R' = R \otimes_k k' \text{ is a complete intersection} \quad \Rightarrow \quad R \text{ is a complete intersection.}$$

Let $\mathfrak{m}$ and $\mathfrak{m}'$ denote the maximal ideals of R and R' respectively.

By assumption, we have $R' = k'[[Y_1, \ldots, Y_r]]/J$, where the ideal J can be generated by r elements. We can assume without loss of generality (by adding extra variables and relations if necessary) that the images of $Y_1, \ldots Y_r$ generate $\mathfrak{m}'$ as a k'-vector space. Now, let

$$\phi : k[[X_1, \ldots, X_r]] \longrightarrow R$$

be a ring homomorphism, such that the images of $X_1, \ldots, X_r$ generate $\mathfrak{m}$ as a k-vector space, and let

$$\phi' : k'[[X_1, \ldots, X_r]] \longrightarrow R'$$

be the extension of scalars. Let I and $I' = I \otimes_k k'$ be the kernels of these two maps. We claim that I' can be generated by r elements. (In fact, this is also true without the assumption that the images of $X_1, \ldots, X_r$ generate $\mathfrak{m}'$ as a k' vector space, although we use this assumption in the proof below. Cf. remark 5.2.) To see this, choose an isomorphism of k'-vector spaces $k'X_1 \oplus \cdots \oplus k'X_r \longrightarrow k'Y_1 \oplus \cdots \oplus k'Y_r$ such that

$$\begin{array}{ccc} k'X_1 \oplus \cdots \oplus k'X_r & \longrightarrow & k'Y_1 \oplus \cdots \oplus k'Y_r \\ \downarrow & & \downarrow \\ \mathfrak{m}' & = & \mathfrak{m}' \end{array}$$

commutes. Extending this map to an isomorphism of k'-algebras,

$$\nu : k'[[X_1, \ldots, X_r]] \longrightarrow k'[[Y_1, \ldots, Y_r]],$$

one sees that $I' = \nu^{-1}(J)$, and hence can be generated by r elements, as claimed.

In particular we have $\dim_{k'}(I'/\mathfrak{m}'I') \leq r$, and

$$\dim_k(I/\mathfrak{m}I) = \dim_{k'}((I/\mathfrak{m}I) \otimes_k k') \leq \dim_{k'}(I'/\mathfrak{m}'I') \leq r.$$

Nakayama's lemma now implies that I can be generated by r elements, and hence R is a complete intersection. This proves the lemma. □

5.10 J-structures

We now turn to the proof of theorem 3.41. In view of the last section we will work in characteristic ℓ. Thus let $\pi : R \twoheadrightarrow T$ be a surjective morphism in the category $\mathcal{C}_k$, where R and T are finite dimensional as k-vector spaces. Let r be a non-negative integer. If $J \lhd k[[S_1, ..., S_r]]$ and $J \subset (S_1, ..., S_r)$ then by a strong J-structure we shall mean a commutative diagram in $\mathcal{C}_k$

$$\begin{array}{ccccc} & & k[[S_1, ..., S_r]] & & \\ & & \downarrow & & \\ k[[X_1, ..., X_r]] & \twoheadrightarrow & R' & \twoheadrightarrow & T' \\ & & \downarrow & & \downarrow \\ & & R & \twoheadrightarrow & T, \end{array}$$

such that

(a) $T'/(S_1, ..., S_r)T' \xrightarrow{\sim} T$ and $R'/(S_1, ..., S_r)R' \twoheadrightarrow R$,

(b) for each ideal $I \supset J$, $I = \ker(k[[S_1, ..., S_r]] \to T'/I)$,

(c) $J = \ker(k[[S_1, ..., S_r]] \to T')$ and $R' \hookrightarrow T' \oplus R$.

We refer to these as *strong* J-structures because we have slightly altered the conditions from section 3.4 for technical convenience in the rest of this section.

We will prove the following result.

Theorem 5.31 *Suppose there exist a sequence of ideals $J_n \lhd k[[S_1, ..., S_r]]$ such that $J_0 = (S_1, ..., S_r)$, $J_n \supset J_{n+1}$, $\bigcap_n J_n = (0)$ and for each n there exists a strong J_n-structure. Then $R \xrightarrow{\sim} T$ and these rings are complete intersections.*

Before proving theorem 5.31 we shall explain how to deduce theorem 3.41 from it. If J is an ideal of $\mathcal{O}[[S_1, ..., S_r]]$ we will use $\bar{J}$ to denote its image in $k[[S_1, ..., S_r]]$. If $\mathcal{S}$ given by

$$\begin{array}{ccccc} & & \mathcal{O}[[S_1, ..., S_r]] & & \\ & & \downarrow & & \\ \mathcal{O}[[X_1, ..., X_r]] & \twoheadrightarrow & R' & \twoheadrightarrow & T' \\ & & \downarrow & & \downarrow \\ & & R & \twoheadrightarrow & T, \end{array}$$

is a J-structure for $R \to T$, let $\mathfrak{a} = \ker(R \longrightarrow T)$, and let R'' denote the ring $\mathrm{Im}\,(R' \to (R/\mathfrak{m}_R\mathfrak{a} \oplus (T'/J)) \otimes_{\mathcal{O}} k)$. Then $\bar{\mathcal{S}}$ given by

$$\begin{array}{ccccc} & & k[[S_1,...,S_r]] & & \\ & & \downarrow & & \\ k[[X_1,...,X_r]] & \twoheadrightarrow & R'' & \twoheadrightarrow & (T'/J) \otimes_{\mathcal{O}} k \\ & & \downarrow & & \downarrow \\ & & (R/\mathfrak{m}_R\mathfrak{a}) \otimes_{\mathcal{O}} k & \twoheadrightarrow & T \otimes_{\mathcal{O}} k, \end{array}$$

is a strong $\bar{J}$-structure for $(R/\mathfrak{m}_R\mathfrak{a}) \otimes_{\mathcal{O}} k \to T \otimes_{\mathcal{O}} k$. If J_n is a sequence of ideals as in theorem 3.41 then $\bar{J}_n$ is a sequence of nested ideals with $\bar{J}_0 = (S_1,...,S_r)$ and $\bigcap_n \bar{J}_n = (0)$. Moreover if for each n there is a J_n-structure $\mathcal{S}_n$ for $R \to T$ then for each n there is a strong $\bar{J}_n$-structure $\bar{\mathcal{S}}_n$ for

$$(R/\mathfrak{m}_R\mathfrak{a}) \otimes_{\mathcal{O}} k \to T \otimes_{\mathcal{O}} k.$$

Then by theorem 5.31 we see that this map is an isomorphism of complete intersections. Lemma 5.29 shows that theorem 3.41 follows. □

Before returning to the proof of theorem 5.31, let us first make some remarks about strong J-structures.

• The set of strong J structures for all ideals $J \subset (S_1,...,S_r) \subset k[[S_1,...,S_r]]$ forms a category, with the obvious notion of morphism.

• If $\mathcal{S}$ is a strong J-structure and if $(S_1,...,S_r) \supset J' \supset J$ then there is a natural J'-structure $\mathcal{S}$ mod J' obtained by replacing T' by T'/J' and R' by the image of $R' \to R \oplus (T'/J')$.

• If R is a finite dimensional k-vector space and if J has finite index in $k[[S_1,...,S_r]]$ then there are only finitely many isomorphism classes of strong J-structure. (This follows because we can bound the order of R' in any J-structure. Explicitly we must have $\#R' \leq (\#R)(\#k[[S_1,...,S_r]]/J)^{\dim_k T}$.)

Lemma 5.32 *Suppose that R is a finite dimensional k vector space. Suppose also that $\{J_n\}$ is a nested (decreasing) sequence of ideals and that $J = \bigcap_n J_n$. If for each n a strong J_n structure exists then a strong J structure exists.*

Proof: We may suppose that each J_n has finite index in $k[[S_1,...,S_r]]$. Let $\mathcal{S}_n$ denote a strong J_n-structure. Let $\mathcal{S}_{n,m} = \mathcal{S}_n$ mod J_m if $m \leq n$. Because there are only finitely many isomorphism classes of strong J_m structure, we may recursively choose integers $n(m)$ such that

- $\mathcal{S}_{n(m),m} \cong \mathcal{S}_{n,m}$ for infinitely many n,
- if $m > 1$ then $\mathcal{S}_{n(m),m-1} \cong \mathcal{S}_{n(m-1),m-1}$.

Let $\mathcal{S}'_m = \mathcal{S}_{n(m),m}$. Then $\mathcal{S}'_m$ is a strong J_m structure and if $m \geq m_1$ then $\mathcal{S}'_m$ mod $J_{m_1} \cong \mathcal{S}'_{m_1}$. One checks that $\mathcal{S} = \varprojlim \mathcal{S}'_m$ is the desired strong J-structure. □

Lemma 5.33 *Suppose that a strong* (0) *structure exists. Then the map* $R \longrightarrow T$ *is an isomorphism, and these rings are complete intersections.*

Proof: Because $k[[S_1, ..., S_r]] \hookrightarrow T'$ (and T' is a finitely generated $k[[S_1, ..., S_r]]$-module by Nakayama's lemma, cf. [Mat], thm. 8.4) we see that the Krull dimension of T' is at least r. On the other hand $k[[X_1, ..., X_r]] \twoheadrightarrow T'$ and so by Krull's principal ideal theorem this map must be an isomorphism. Thus

$$k[[X_1, ..., X_r]] \xrightarrow{\sim} R' \xrightarrow{\sim} T'.$$

Hence we have that

$$k[[X_1, ..., X_r]]/(S_1, ..., S_r) \xrightarrow{\sim} R'/(S_1, ..., S_r) \xrightarrow{\sim} T,$$

and the lemma follows. □

Theorem 5.31 follows at once from these two lemmas. □

Bibliography

[Ant2] W. Kuyk, P. Deligne, eds., Modular Functions of One Variable II, Lecture Notes in Math. **349**, Springer-Verlag, New York, Berlin, Heidelberg, 1973.

[Ant3] W. Kuyk, J.-P. Serre, eds., Modular Functions of One Variable III, Lecture Notes in Math. **350**, Springer-Verlag, New York, Berlin, Heidelberg, 1973.

[Ant4] B. Birch, W. Kuyk, eds., Modular Functions of One Variable IV, Lecture Notes in Math. **476**, Springer-Verlag, New York, Berlin, Heidelberg, 1975.

[Ant5] J.-P. Serre, D. Zagier, eds., Modular Functions of One Variable V, Lecture Notes in Math. **601**, Springer-Verlag, New York, Berlin, Heidelberg, 1977.

[AL] A.O.L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160.

[BCEM] J. Buhler, R. Crandall, R. Ernvall, T. Metsänkylä, *Irregular primes and cyclotomic invariants to four million*, Math. Comp. **61** (1993), 151–153.

[BK] S. Bloch, K. Kato, *L-functions and Tamagawa numbers of motives*, in: The Grothendieck Festschrift I, Progress in Math. **86**, Birkhäuser, Boston, Basel, Berlin, 1990, pp. 333–400.

[BLR] N. Boston, H. Lenstra, K. Ribet, *Quotients of group rings arising from two-dimensional representations*, C. R. Acad. Sc. **t312** (1991), 323–328.

[BM] N. Boston, B. Mazur, *Explicit universal deformations of Galois representations*, Algebraic Number Theory, 1–21, Adv. Stud. Pure Math. **17**, Acad. Press, Boston MA, 1989.

[Bour] N. Bourbaki, Algèbre (Éléments de Mathématiques, Fasc. XXIII), Hermann, Paris, 1958.

[Ca1] H. Carayol, *Sur les représentations p-adiques associées aux formes modulaires de Hilbert*, Ann. Sci. Ec. Norm. Super. **19** (1986), 409–468.

[Ca2] H. Carayol, *Sur les représentations galoisiennes modulo ℓ attachées aux formes modulaires*, Duke Math. J. **59** (1989), 785–801.

[Ca3] H. Carayol, *Formes modulaires et représentations galoisiennes à valeurs dans un anneau local complet*, in: p-adic Monodromy and the Birch-Swinnerton-Dyer Conjecture (B. Mazur and G. Stevens, eds.), Contemporary Math. **165**, 1994, pp. 213–237.

[Co] J. Coates, S.T. Yau, eds., Elliptic Curves, Modular Forms and Fermat's Last Theorem, International Press, Cambridge, 1995.

[CPS] E. Cline, B. Parshall, L. Scott, *Cohomology of finite groups of Lie type I*, Publ. Math. IHES **45** (1975), 169–191.

[Cr] J. Cremona, Algorithms for Modular Elliptic Curves, Cambridge Univ. Press, Cambridge, 1992.

[CR] C.W. Curtis, I. Reiner, Representation Theory of Finite Groups and Associative Algebras, Wiley Interscience, New York, 1962.

[CS] G. Cornell and J.H. Silverman, eds., Arithmetic Geometry, Springer-Verlag, New York, Berlin, Heidelberg, 1986.

[CW] J. Coates, A. Wiles, *On the conjecture of Birch and Swinnerton-Dyer*, Inv. Math. **39** (1977), 223–251.

[Da1] H. Darmon, *Serre's conjectures*, in [Mu2], pp. 135–153.

[Da2] H. Darmon, *The Shimura-Taniyama Conjecture, (d'après Wiles)*, (in Russian) Uspekhi Mat. Nauk 50 (1995), no. 3(303), 33–82. (English version to appear in Russian Math Surveys)

[De] P. Deligne, *Formes modulaires et représentations ℓ-adiques*, in: Lecture Notes in Math. **179**, Springer-Verlag, New York, Berlin, Heidelberg, 1971, pp. 139–172.

[DR] P. Deligne, M. Rapoport, *Les schémas de modules de courbes elliptiques*, in [Ant2], pp. 143–316.

[dS] E. de Shalit, *On certain Galois representations related to the modular curve $X_1(p)$*, to appear in Compositio Math.

[dSL] B. de Smit, H.W. Lenstra, *Explicit construction of universal deformation rings*, to appear in the Proceedings of the Conference on Fermat's Last Theorem, Boston University, August 9–18, 1995.

[DS] P. Deligne, J.-P. Serre, *Formes modulaires de poids* 1, Ann. Sci. Ec. Norm. Sup. **7** (1974), 507–530.

[Di1] F. Diamond, *The refined conjecture of Serre*, in [Co], pp. 22–37.

[Di2] F. Diamond, *On deformation rings and Hecke rings*, to appear in Annals of Math.

[Dic1] L.E. Dickson, History of the Theory of Numbers, Vol. II, Chelsea Publ. Co., New York, 1971.

[Dic2] L.E. Dickson, Linear Groups with an Exposition of the Galois Field Theory, Teubner, Leipzig, 1901.

[Dir] P.G.L. Dirichlet, *Mémoire sur l'impossibilité de quelques équations indéterminées du cinquième degré*, Jour. für Math. (Crelle) **3** (1828), 354–375.

[DI] F. Diamond and J. Im, *Modular forms and modular curves*, in [Mu2], pp. 39–133.

[DK] F. Diamond and K. Kramer, *Modularity of a family of elliptic curves* Math. Research Letters **2** (1995), 299–305.

[DO] K. Doi, M. Ohta, *On some congruences between cusp forms on* $\Gamma_0(N)$, in [Ant5], pp. 91–106.

[Dr] V.G. Drinfeld, *Elliptic modules*, (Russian) Math Sbornik **94** (1974), 594–627. (English translation: Math. USSR, Sbornik **23** (1973).)

[DT1] F. Diamond and R. Taylor, *Non-optimal levels of mod ℓ modular representations*, Invent. Math. **115** (1994) 435–462.

[DT2] F. Diamond and R. Taylor, *Lifting modular mod ℓ representations*, Duke Math. J. **74** (1994) 253–269.

[Edi] B. Edixhoven, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992), 563–594.

[Edw] H.M. Edwards, Fermat's Last Theorem: A genetic introduction to algebraic number theory, Graduate Texts in Math. **50**, Springer-Verlag, New York, Berlin, Heidelberg, 1977.

[Ei] M. Eichler, *Quaternäre quadratische Formen und die Riemannsche Vermutung für die Kongruenzzetafunktion*, Arch. Math. **5** (1954) 355–366.

[Fa] G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Inv. Math. **73** (1983), 349–366. (English translation in [CS].)

[Fl1] M. Flach, *A finiteness theorem for the symmetric square of an elliptic curve*, Inv. Math. **109** (1992), 307–327.

[Fl2] M. Flach, *On the degree of modular parametrizations*, Séminaire de théorie des nombres, Paris, 1991-92, 23-36; Prog. in Math. **116**, Birkhäuser, Boston, MA 1993.

[FL] J.-M. Fontaine and G. Laffaille, *Construction de représentations p-adiques*, Ann. Sci. Ec. Norm. Super. **15** (1982), 547–608.

[FM] J.-M. Fontaine and B. Mazur, *Geometric Galois representations*, in [Co], 41–78.

[Fo1] J.-M. Fontaine, *Groupes finis commutatifs sur les vecteurs de Witt*, C. R. Acad. Sc. **280** (1975), 1423–1425.

[Fo2] J.-M. Fontaine, *Groupes p-divisibles sur les corps locaux*, Astérisque **47–48** (1977).

[For] O. Forster, Lecures on Riemann Surfaces, Springer-Verlag, New York, Berlin, Heidelberg, 1981.

[Fr] G. Frey, *Links between solutions of $A - B = C$ and elliptic curves*, in: Number Theory, Ulm 1987, Proceedings, Lecture Notes in Math. **1380**, Springer-Verlag, New York, Berlin, Heidelberg, 1989, pp. 31–62.

[Gra] A. Granville, *The set of exponents for which Fermat's Last Theorem is true, has density one*, Comptes Rendus de l'Académie des Sciences du Canada **7** (1985), 55-60.

[Gre] R. Greenberg, *Iwasawa theory for p-adic representations*, in: Algebraic Number Theory, in honor of K. Iwasawa, Adv. Studies in Pure Math. **17**, Academic Press, San Diego, 1989, pp. 97–137.

[Gro] B.H. Gross, *A tameness criterion for Galois representations associated to modular forms* mod$p$, Duke Math. J. **61** (1990), 445–517.

[Ha] R. Hartshorne, Algebraic Geometry, Graduate Texts in Math. **52**, Springer-Verlag, New York, Berlin, Heidelberg, 1977.

[HB] D.R. Heath-Brown, *Fermat's Last Theorem for "almost all" exponents*, Bull. London. Math. Soc. 17 (1985), 15-16

[He] Y. Hellegouarch, *Points d'ordre $2p^h$ sur les courbes elliptiques*, Acta Arith. **26** (1974/75) 253–263.

[Hi1] H. Hida, *Congruences of cusp forms and special values of their zeta functions*, Inv. Math. **63** (1981), 225–261.

[Hi2] H. Hida, *A p-adic measure attached to the zeta functions associated with two elliptic modular forms. I*, Inv. Math. **79** (1985), 159–195.

[Hi3] H. Hida, *Galois representations into $GL_2(\mathbb{Z}_p[[X]])$ attached to ordinary cusp forms,* Inv. Math. **85** (1986), 545–613.

[Hu] B. Huppert, Endliche Gruppen I, Grundlehren Math. Wiss. **134**, Springer-Verlag, New York, Berlin, Heidelberg, 1983.

[Ig] J. Igusa, *Kroneckerian model of fields of elliptic modular functions*, Amer. J. Math. **81**, 561–577 (1959).

[Ih] Y. Ihara, *On modular curves over finite fields*, in: Proc. Intl. Coll. on Discrete Subgroups of Lie Groups and Applications to Moduli, Bombay, 1973, pp. 161–202.

[Kam] S. Kamienny, *Torsion points on elliptic curves over fields of higher degree*, Inter. Math. Res. Not. **6**, 1992, 129–133.

[Kat] N.M. Katz, *p-adic properties of modular schemes and modular forms*, in [Ant3], pp. 70–189.

[Kh] C. Khare, *Congruences between cusp forms: the (p,p) case*, to appear in Duke Math. J.

[Ki] F. Kirwan, Complex Algebraic Curves, LMS Student Texts **23**, Cambridge Univ. Press, Cambridge, 1993.

[KL] D.S. Kubert, S. Lang, Modular Units. Springer-Verlag, 1981.

[KM] N.M. Katz, and B. Mazur, Arithmetic Moduli of Elliptic Curves, Annals of Math. Studies **108**, Princeton Univ. Press, Princeton, 1985.

[Kn] A.W. Knapp, Elliptic Curves, Princeton Univ. Press, Princeton, 1992.

[Koc] H. Koch, Galoissche Theorie der p-Erweiterungen, Springer-Verlag, New York, Berlin, Heidelberg, 1970.

[Kol] V.A. Kolyvagin, *Finiteness of $E(\mathbb{Q})$ and $sha(E/\mathbb{Q})$ for a subclass of Weil curves*, Izv. Akad. Nauk. SSSR Ser. Mat. **52** (3) (1988), 522–540. (English translation: Math. USSR, Izvestiya **32** (1989).)

[Ku] E. Kunz, Introduction to Commutative Algebra and Algebraic Geometry, Birkhäuser, Boston, Basel, Berlin, 1985.

[La1] S. Lang, Algebraic Number Theory, Addison-Wesley, Reading, MA 1970.

[La2] S. Lang, Introduction to Modular Forms, Grundlehren Math. Wiss. **222**, Springer-Verlag, New York, Berlin, Heidelberg, 1976.

[Leg] A.M. Legendre, *Sur quelques objets d'analyse indéterminée et particulièrement sur le théorème de Fermat*, Mém. Acad. R. Sc. de l'Institut de France **6**, Paris, 1827.

[Len] H.W. Lenstra, *Complete intersections and Gorenstein rings*, in [Co], pp. 99–109.

[Liv] R. Livné, *On the conductors of mod ℓ representations coming from modular forms*, J. Number Theory **31** (1989), 133–141.

[Ll1] R.P. Langlands, *Modular forms and ℓ-adic representations*, in [Ant2], pp. 361–500.

[Ll2] R.P. Langlands, Base Change for $GL(2)$, Annals of Math. Studies **96**, Princeton Univ. Press, Princeton, 1980.

[LO] S. Ling, J. Oesterlé, *The Shimura subgroup of $J_0(N)$*, Astérisque **196–197** (1991), 171–203.

[Mat] H. Matsumura, Commutative Ring Theory, Cambridge Studies in Adv. Math. **8**, Cambridge Univ. Press, Cambridge, 1986.

[Maz1] B. Mazur, *Modular curves and the Eisenstein ideal*, Publ. Math. IHES **47** (1977), 33–186.

[Maz2] B. Mazur, *Rational isogenies of prime degree*, Inv. Math. **44** (1978), 129–162.

[Maz3] B. Mazur, *Deforming Galois representations*, in: Galois groups over $\mathbb{Q}$, Y. Ihara, K. Ribet, J-P. Serre, eds., MSRI Publ. **16**, Springer-Verlag, New York, Berlin, Heidelberg, 1989, pp. 385–437.

[Mc] W. McCallum, *On the method of Coleman and Chabauty*, Math. Ann. **299** (1994), 565–596.

[Mer] L. Merel, *Bornes pour la torsion des courbes elliptiques sur les corps de nombres*, Inventiones Math., to appear.

[Mes] J-F. Mestre, *Constructions polynomiales et théorie de Galois*, Proceedings of the ICM, Zurich, August 1994.

[Mi] J.S. Milne, Arithmetic Duality Theorems, Perspectives in Math., Academic Press, San Diego, 1986.

[MRi] B. Mazur, K. Ribet, *Two-dimensional representations in the arithmetic of modular curves*, Astérisque **196–197** (1991), 215–255.

[MRo] B. Mazur, L. Roberts, *Local Euler characteristics*, Inv. Math. **9** (1970), 201–234.

[MT] B. Mazur, J. Tilouine, *Représentations galoisiennes, différentielles de Kähler et "conjectures principales"*, Publ. Math. IHES **71** (1990), 65–103.

[Mu1] V.K. Murty, Introduction to Abelian Varieties, CRM Monograph Series **3**, AMS Publ., Providence, 1993.

[Mu2] V.K. Murty, ed., Seminar on Fermat's Last Theorem, CMS Conference Proceedings **17**, AMS Publ., Providence, 1995.

[MW] B. Mazur, A. Wiles, *Class fields of abelian extensions of* $\mathbb{Q}$, Inv. Math. **76** (1984), 179–330.

[Na] K. Nagao, *An example of an elliptic curve over* $\mathbb{Q}(T)$ *with rank* ≥ 13, Proc. of The Japan Acad., **70**, 1994, 152-153.

[Oda] T. Oda, *The first De Rham cohomology group and Dieudonné modules*, Ann. Sci. Ec. Norm. Super. **2** (1969), 63–135.

[Oe] J. Oesterlé, *Nouvelles aproches du "théorème" de Fermat"*, Séminaire Bourbaki no. 694 (1987-88), Astérisque **161–162** (1988), 165–186.

[Ogg] A. Ogg, Modular Forms and Dirichlet Series, Benjamin, New York, 1969.

[Oo] F. Oort, Commutative Group Schemes, Lecture Notes in Math. **15**, Springer-Verlag, New York, Berlin, Heidelberg, 1966.

[Ram] R. Ramakrishna, *On a variation of Mazur's deformation functor*, Compositio Math. **87** (1993), 269–286.

[Ray] M. Raynaud, *Schémas en groupes de type* $(p, p, ..., p)$, Bull. Soc. Math. France **102** (1974), 241–280.

[R1] K. Ribet, *A modular construction of unramified p-extensions of* $\mathbb{Q}(\mu_p)$, Inv. Math. **34**, 151-162.

[R2] K. Ribet, *Twists of modular forms and endomorphisms of abelian varieties*, Math. Ann. **253** (1980), 43-62.

[R3] K.A. Ribet, *The* ℓ*-adic representations attached to an eigenform with Nebentypus: a survey*, in [Ant5], pp. 17–52.

[R4] K.A. Ribet, *Congruence relations between modular forms*, Proc. ICM, 1983, 503–514.

[R5] K.A. Ribet, *On modular representations of* $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ *arising from modular forms*, Inv. Math. **100** (1990), 431–476.

[R6] K.A. Ribet, *Report on mod ℓ representations of* $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, in: Motives, Proc. Symp. in Pure Math. **55** (2), 1994, pp. 639–676.

[R7] K.A. Ribet, *Abelian varieties over* $\mathbb{Q}$ *and modular forms*, 1992 Proceedings of KAIST Mathematics Workshop, Korea Advanced Institute of Science and Technology, Taejon, 1992, pp. 53–79.

[R8] K.A. Ribet, *Galois representations and modular forms*, Bull. AMS. **32** (1995), 375–402.

[Ri] P. Ribenboim, 13 Lectures on Fermat's Last Theorem, Springer-Verlag, New York, Berlin, Heidelberg, 1979.

[Sa] T. Saito, *Conductor, discriminant and the Noether formula for arithmetic surfaces*, Duke Math. J. **57** (1988), 151–173.

[Se1] J.-P. Serre, Cohomologie Galoisienne, Lecture Notes in Math. **5**, Springer-Verlag, New York, Berlin, Heidelberg, 1964.

[Se2] J.-P. Serre, Corps Locaux, Hermann, Paris, 1962.

[Se3] J.-P. Serre, Algèbre Locale, Multiplicités, Cours au Collège de France, 1957-58, Lecture Notes in Math. **11**, Springer-Verlag, New York, Berlin, Heidelberg, 1965.

[Se4] J.-P. Serre, A Course in Arithmetic, Graduate Texts in Math. **7**, Springer-Verlag, New York, Berlin, Heidelberg, 1973.

[Se5] J.-P. Serre, Abelian ℓ-adic Representations and Elliptic Curves, Benjamin, New York, 1968.

[Se6] J.-P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Inv. Math. **15** (1972), 259–331.

[Se7] J.-P. Serre, *Sur les représentations modulaires de degré* 2 *de* $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, Duke Math. J. **54** (1987), 179–230.

[Sha] S. Shatz, *Group schemes, formal groups, and p-divisible groups*, in [CS], pp. 29–78.

[Shi1] G. Shimura, *Correspondances modulaires et les fonctions ζ de courbes algébriques*, J. Math. Soc. Japan **10** (1958), 1–28.

[Shi2] G. Shimura, Introduction to the Arithmetic Theory of Automorphic Functions, Princeton Univ. Press, Princeton, 1971.

[Shi3] G. Shimura, *On the factors of the jacobian variety of a modular function field*, J. Math. Soc. Japan **25** (1973), 523–544.

[Shi4] G. Shimura, *On the holomorphy of certain Dirichlet series*, Proc. LMS (3) **31** (1975), 79–98.

[Shi5] G. Shimura, *The special values of the zeta functions associated with cusp forms*, Comm. Pure Appl. Math. **29** (1976), 783–804.

[Si1] J. Silverman, The Arithmetic of Elliptic Curves, Graduate Texts in Math. **106**, Springer-Verlag, New York, Berlin, Heidelberg, 1986.

[Si2] J. Silverman, Advanced Topics in the Arithmetic of Elliptic Curves, Graduate Texts in Math. **151**, Springer-Verlag, New York, Berlin, Heidelberg, 1994.

[St] J. Sturm, *Special values of zeta functions, and Eisenstein series of half integral weight*, Amer. J. Math. **102** (1980), 219–240.

[ST] J.-P. Serre, J.T. Tate, *Good reduction of abelian varieties*, Annals of Math. **88** (1968), 492–517.

[Su] J. Suzuki, *On the generalized Wieferich criterion*, Proc. Japan Acad. **70** (1994), 230-234.

[Ta] J.T. Tate, *Algorithm for determining the type of a singular fiber in an elliptic pencil*, in [Ant4], pp. 33–52.

[Te] G. Terjanian, *Sur l'équation* $x^{2p} + y^{2p} = z^{2p}$, C.R. Acad. Sci. Paris, **285** (1977) 973-975.

[Th] F. Thaine, *On the ideal class groups of real abelian number fields*, Annals of Math. **128** (1988), 1–18.

[Tu] J. Tunnell, *Artin's conjecture for representations of octahedral type*, Bull. AMS **5** (1981), 173–175.

[TW] R. Taylor, A. Wiles, *Ring theoretic properties of certain Hecke algebras*, Annals of Math. **141** (1995), 553–572.

[Wa] L. Washington, Introduction to Cyclotomic Fields, Graduate Texts in Math. **83**, Springer-Verlag, New York, Berlin, Heidelberg, 1982.

[We1] A. Weil, Variétés Abéliennes et Courbes Algébriques, Hermann, Paris, 1948.

[We2] A. Weil, *Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Math. Ann. **168** (1967), 165–172.

[W1] A. Wiles, *On p-adic representations for totally real fields*, Annals of Math. **123** (1986), 407–456.

[W2] A. Wiles, *On ordinary* λ*-adic representations associated to modular forms*, Inv. Math. **94** (1988), 529–573.

[W3] A. Wiles, *Modular elliptic curves and Fermat's Last Theorem*, Annals of Math. **141** (1995), 443–551.